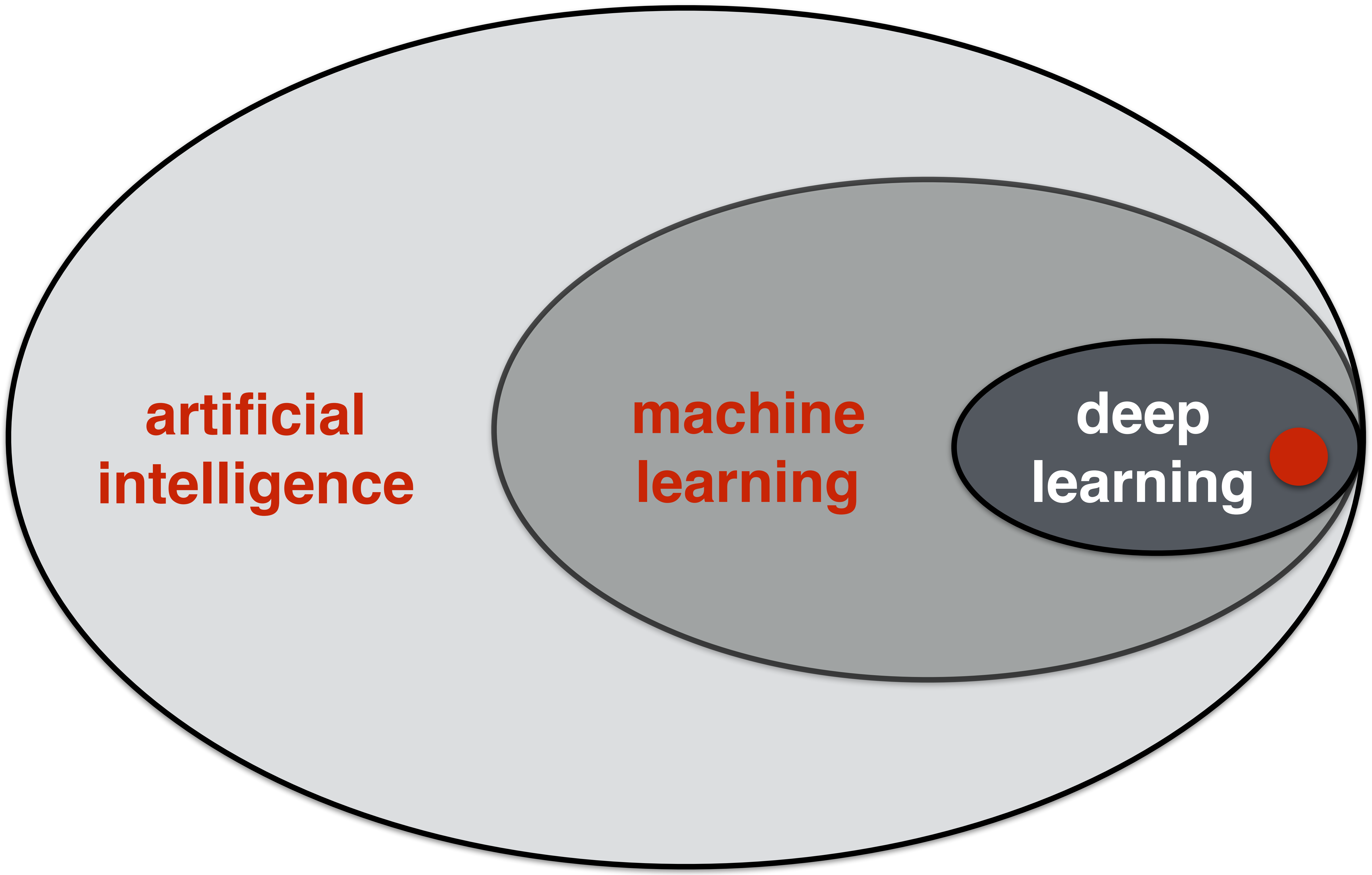
Deep Learning for Computational Scientists

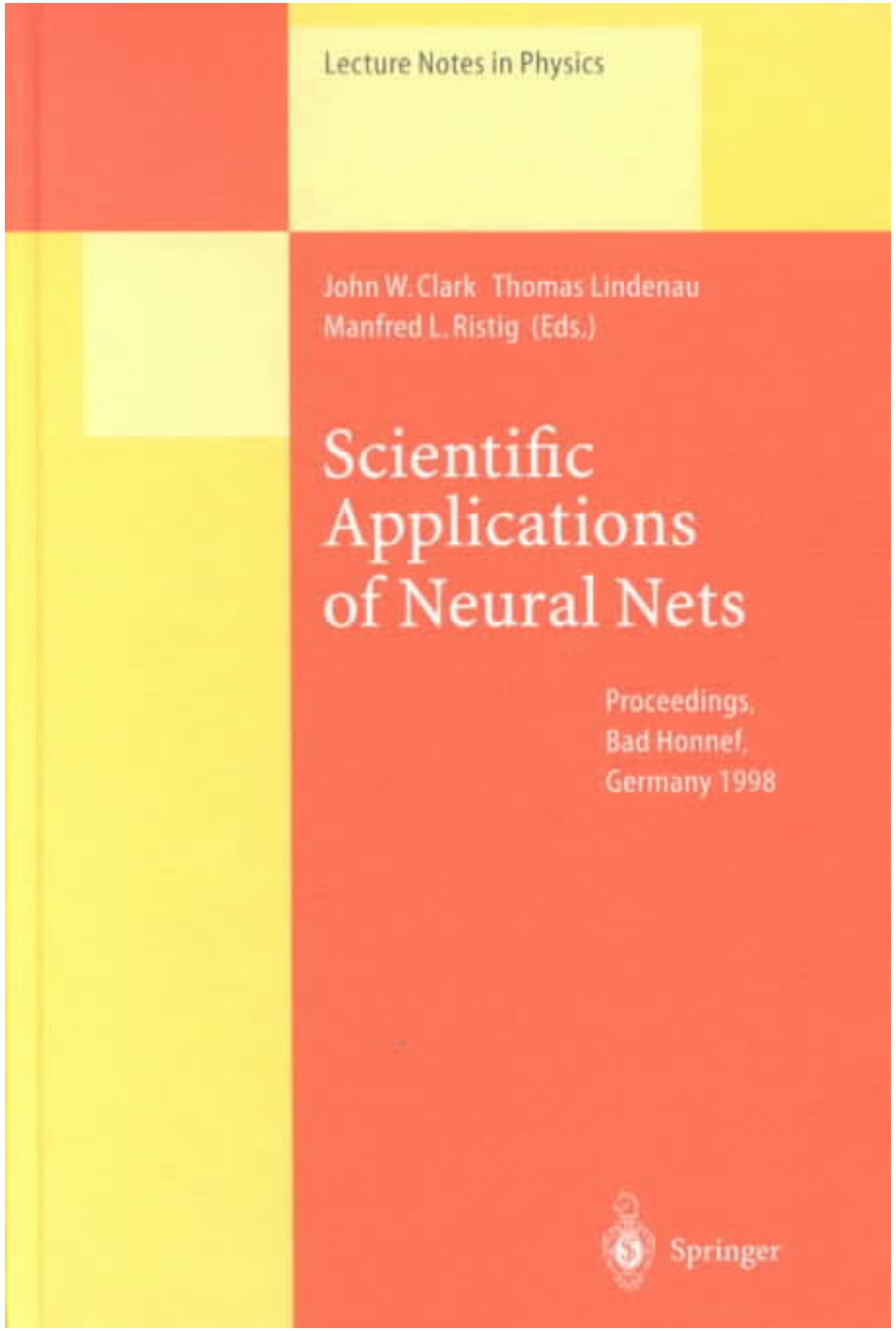
Lei Wang (王磊)
https://wangleiphy.github.io

Institute of Physics, Beijing
Chinese Academy of Sciences

Why deep learning ?



Game changing technology for scientific research especially computational science



Gem in between this and last hype cycles

8 Doing Science With Neural Nets: Pride and Prejudice

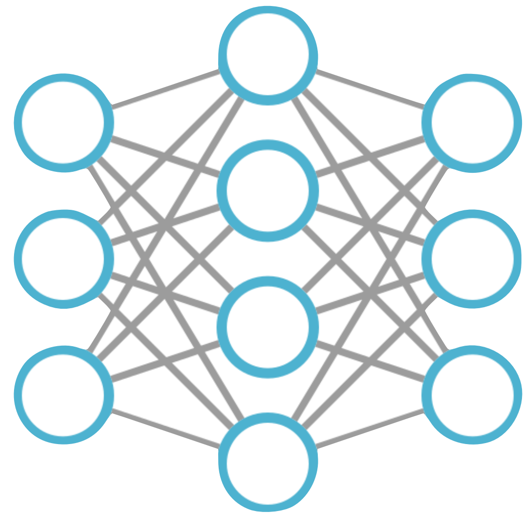
When neural networks re-emerged on the scene in the mid-80s as a new and glamorous computational paradigm, the initial reaction in some sectors of the scientific community was perhaps too enthusiastic and not sufficiently critical. There was a tendency on the part of practitioners to oversell the powers of neural-network or "connectionist" solutions relative to conventional techniques – where conventional techniques can include both traditional theory-rich modeling and established statistical methods. The last five years have seen a correction phase, as some of the practical limitations of neural-network approaches have become apparent, and as scientists have become better acquainted with the wide array of advanced statistical tools that are currently available.

Why now, again ?

What has changed ?

What has not ?

Plan



Hitchhiker's guide to deep learning

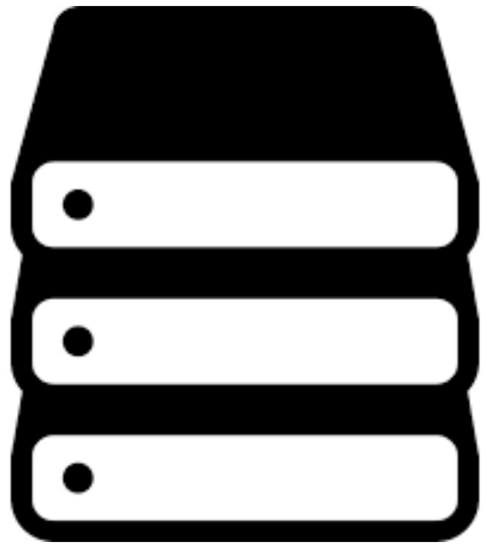
Secrets behind deep learning

Hands on time

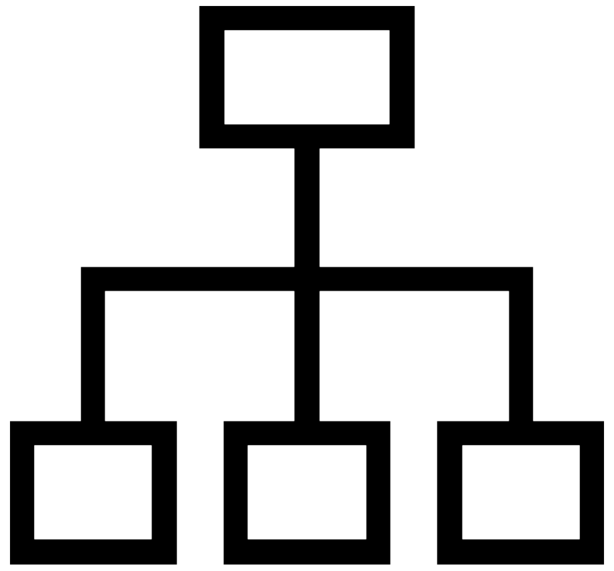
Don't Panic!

Key components

Data



Model



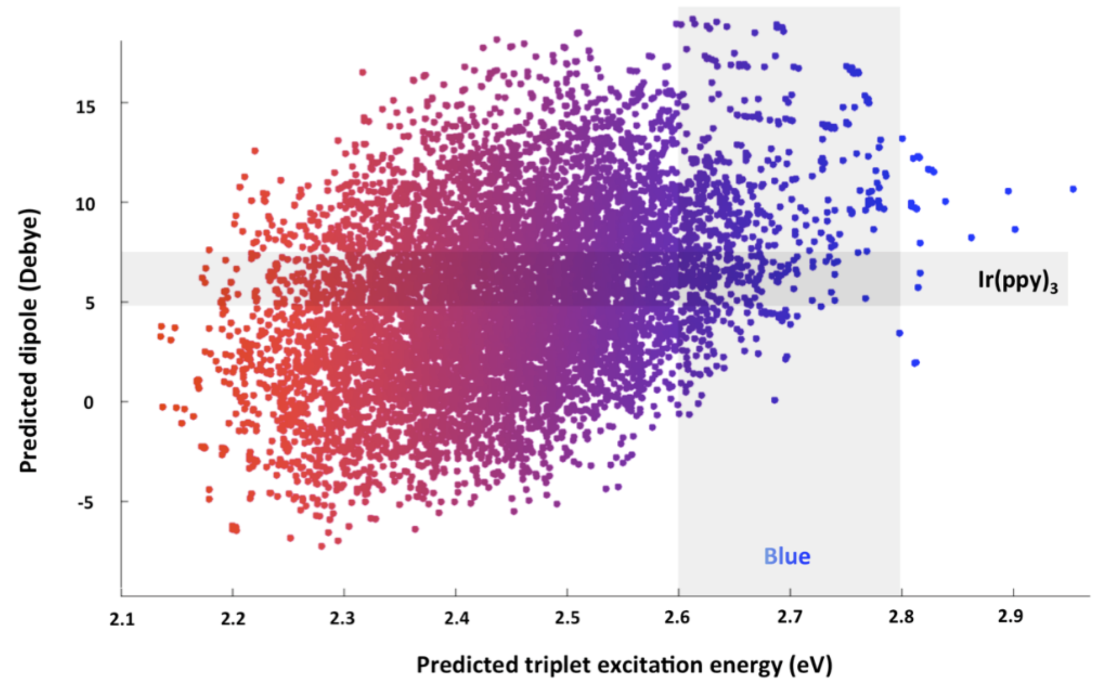
Cost function

Optimization

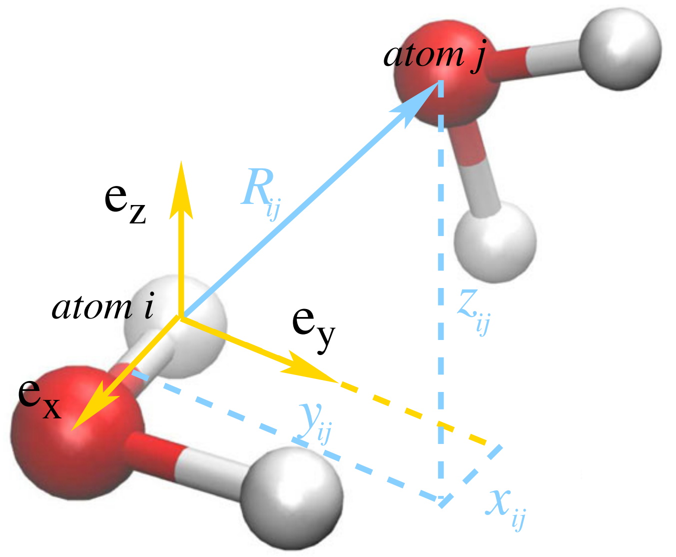
∂

Switch to blackboard

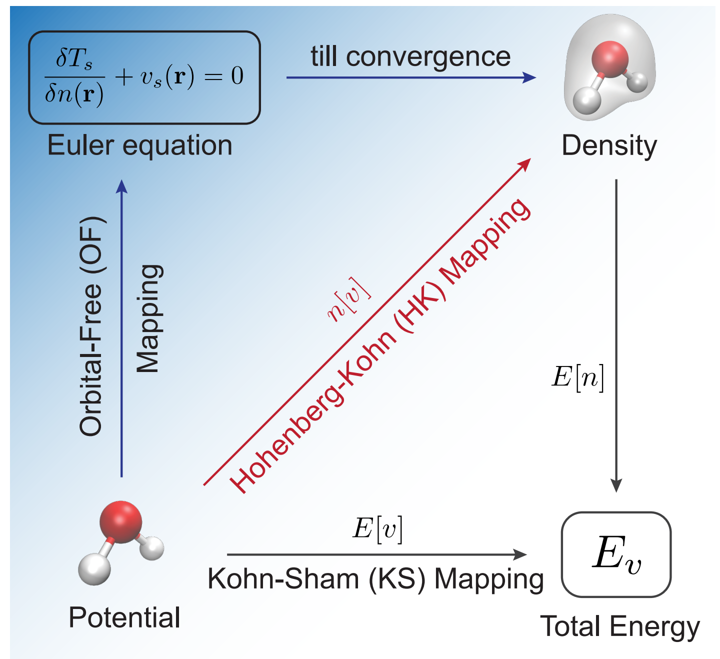
Some applications



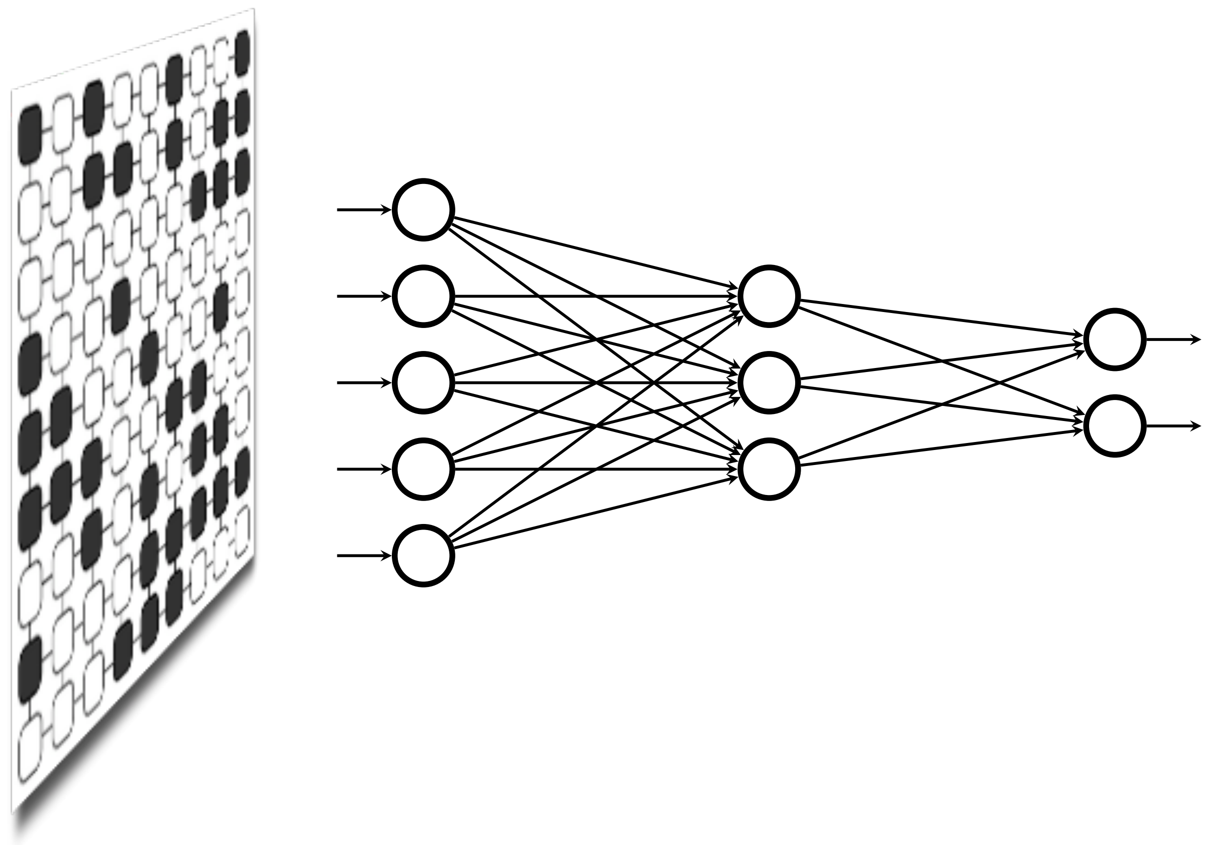
Materials informatics



Molecular simulation

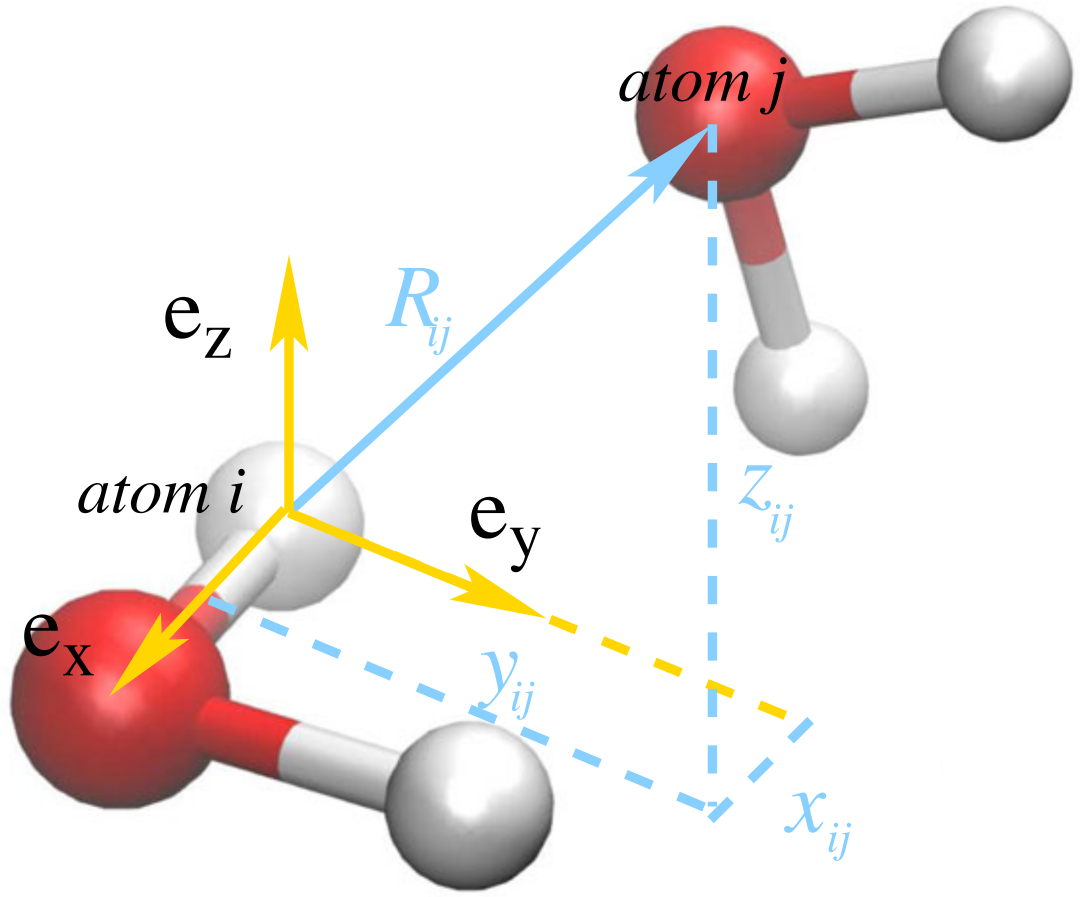


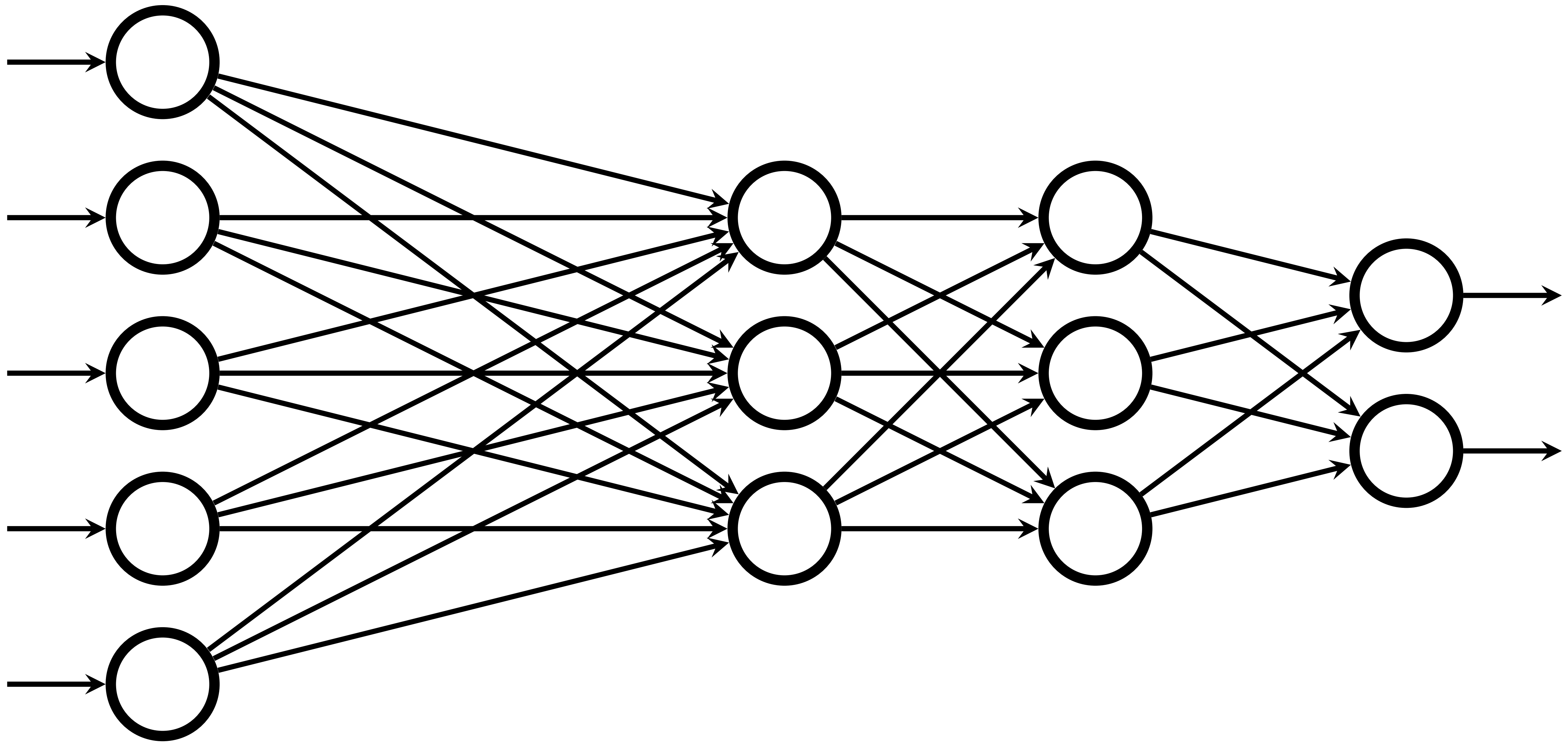
Density functionals



“Phase” recognition

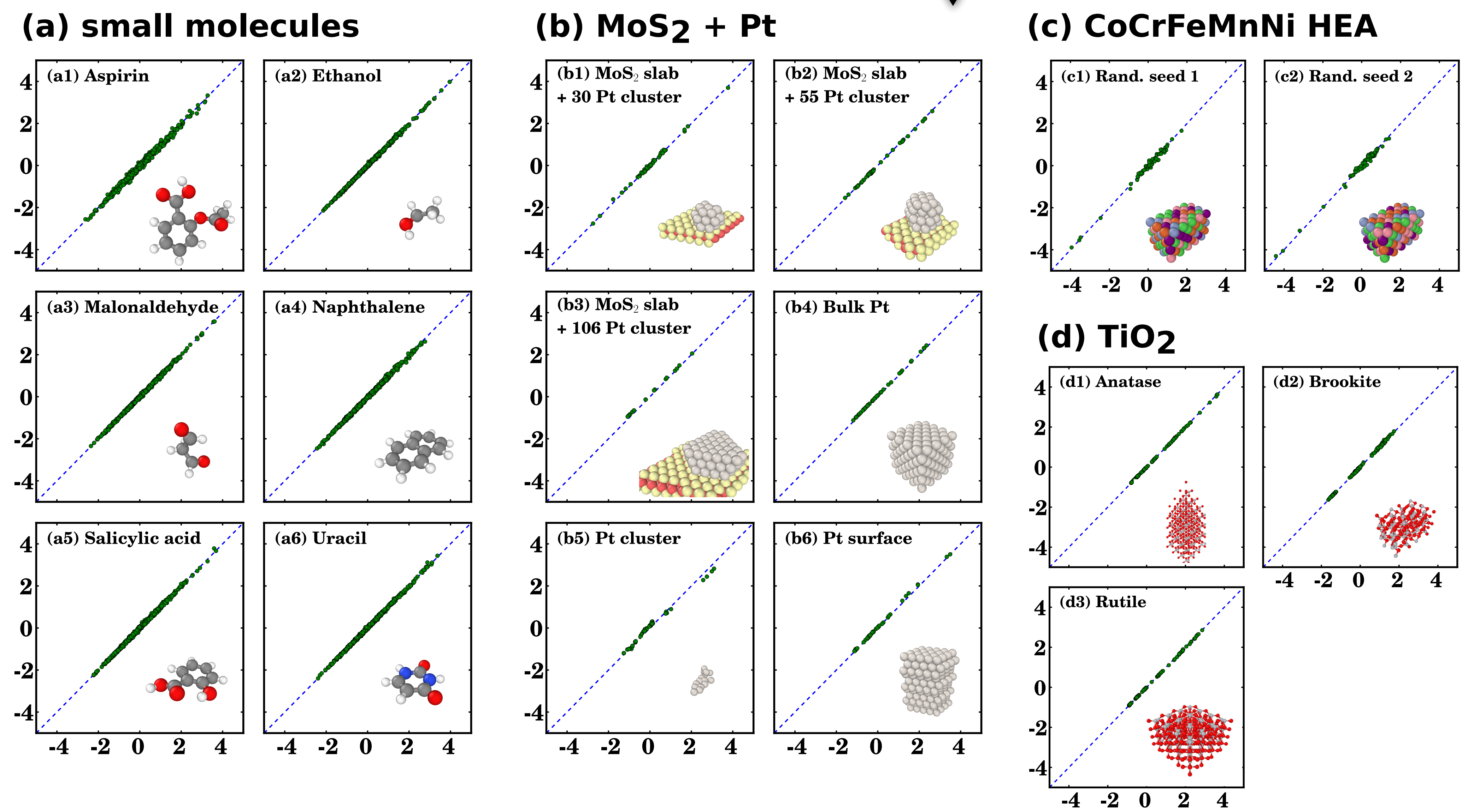
Machine learning energy potential





energy, force…

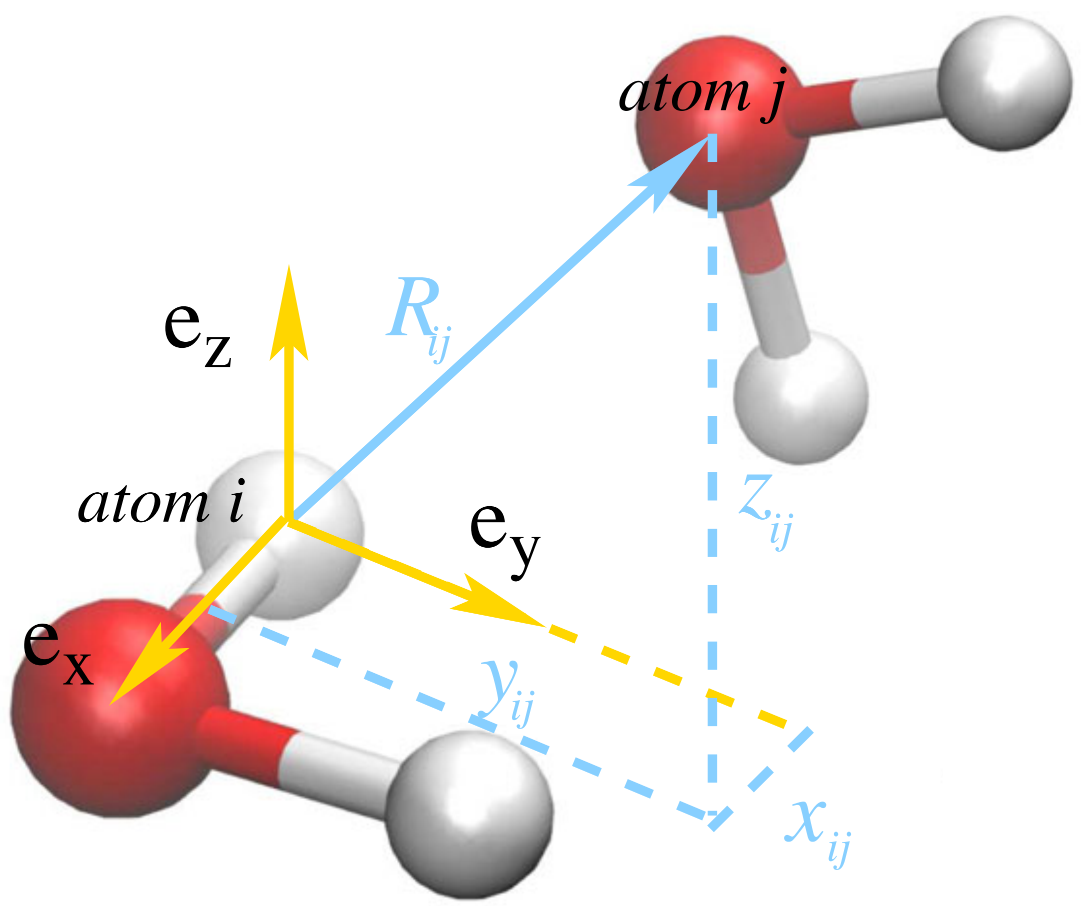
Atom species, position…



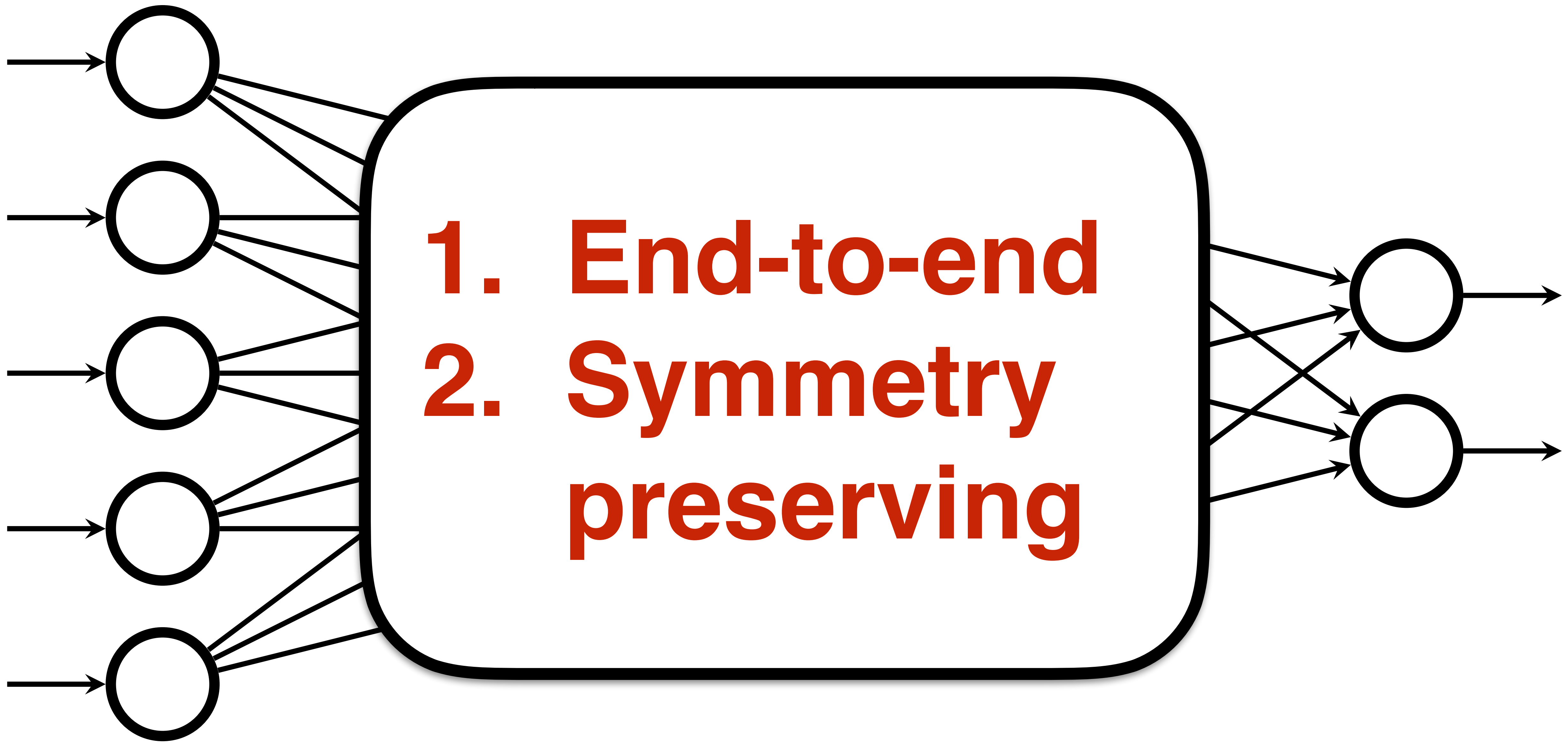
Zhang, Han, Wang, Car, E, PRL 2018

Zhang, Han, Wang, Saidi, Car, E, NIPS 2018

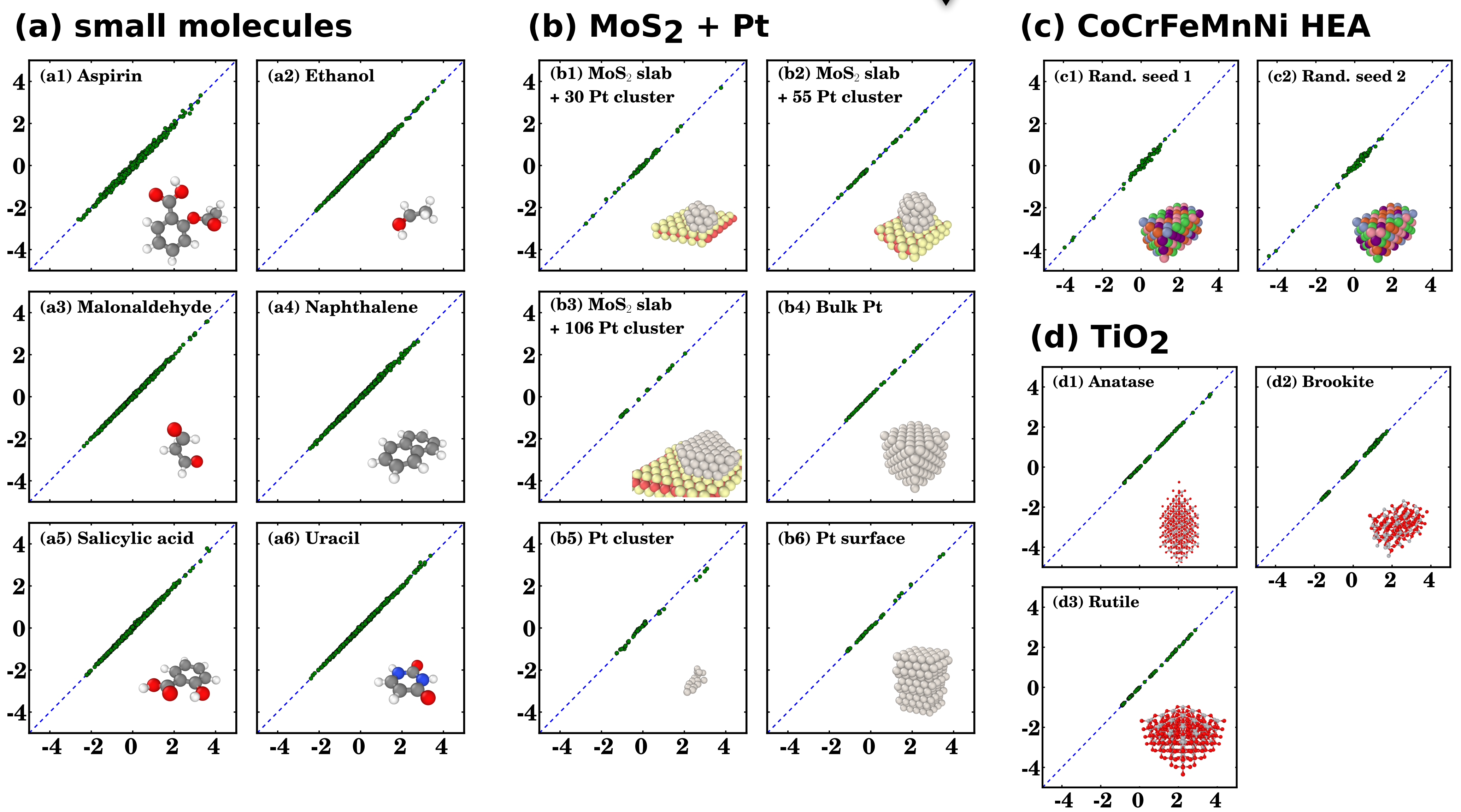
Machine learning energy potential



Atom species, position...



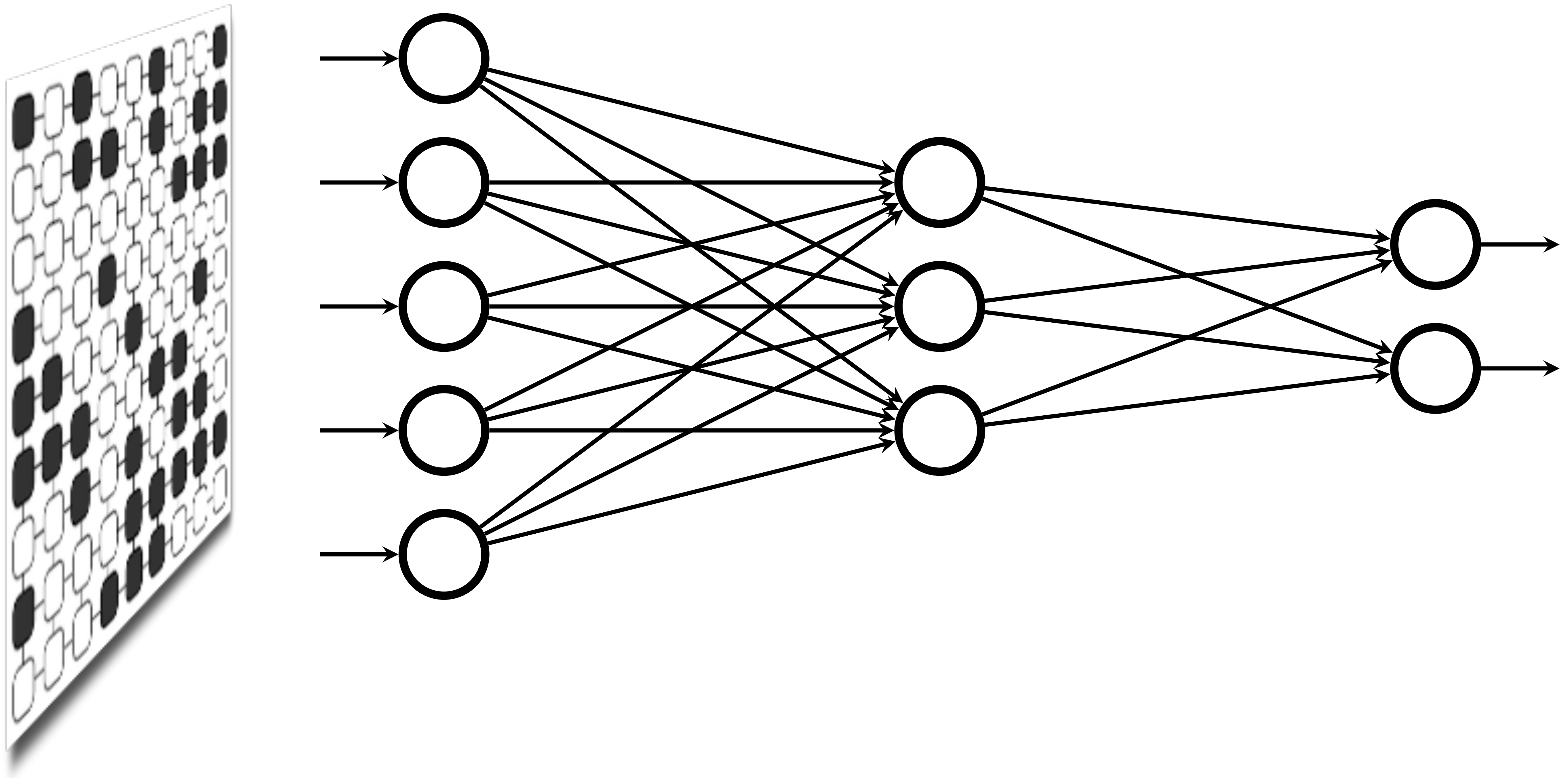
energy, force...



Zhang, Han, Wang, Car, E, PRL 2018

Zhang, Han, Wang, Saidi, Car, E, NIPS 2018

Phase classifications



"Machine Learning Phase of Matter"

Carrasquilla and Melko, 1605.01735

+ many more on quantum spins, fermions, disordered, topological systems …

Deep learning is more than function fitting

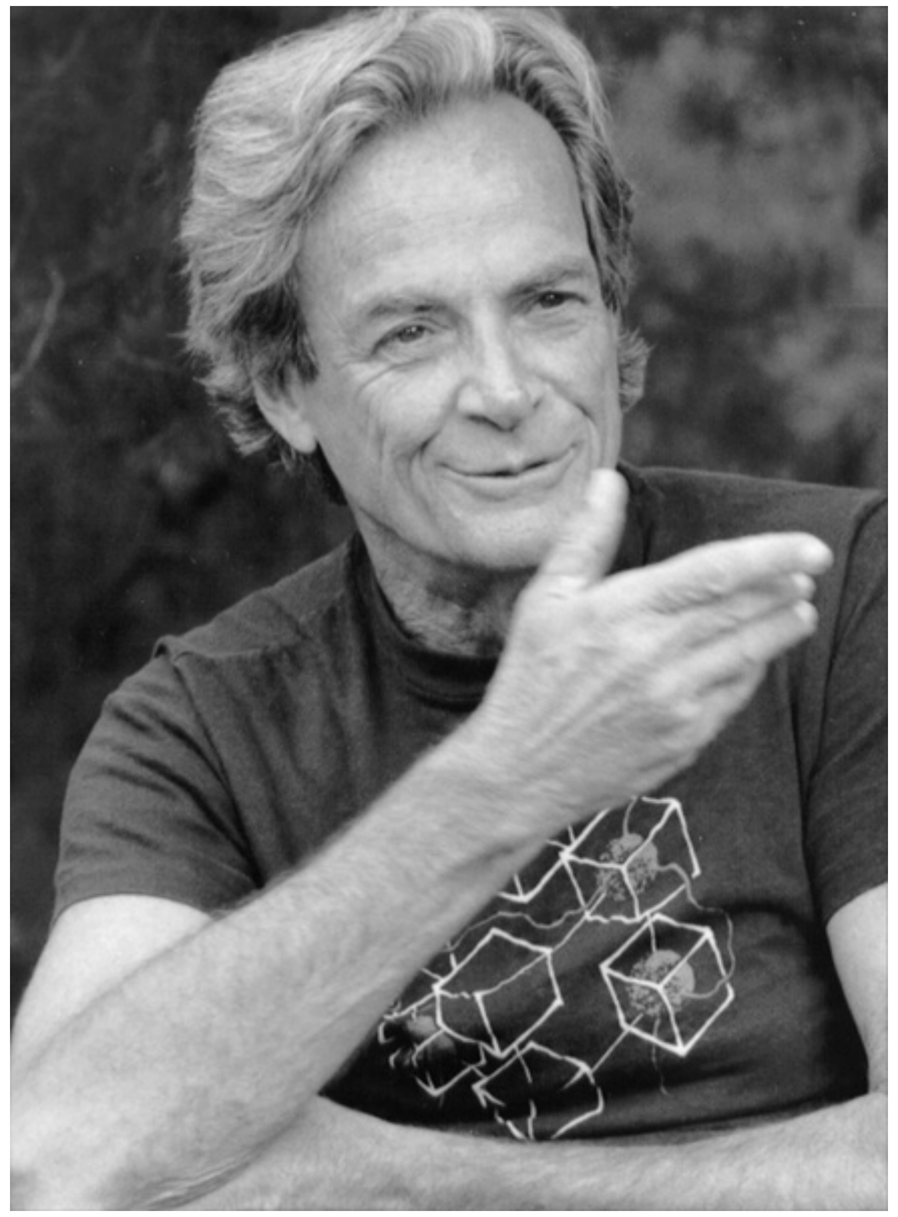
Discriminative

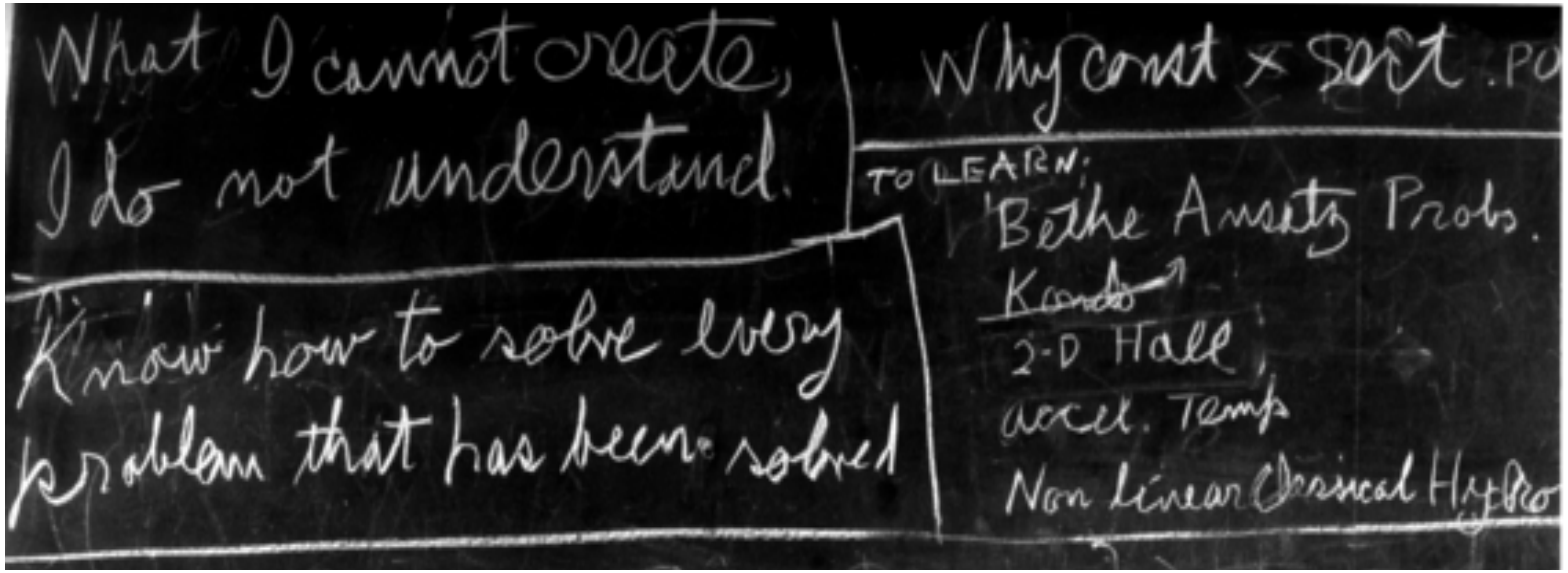
$y = f(\boldsymbol{x})$ or $p(y \mid \boldsymbol{x})$

Generative

$p(\boldsymbol{x}, y)$

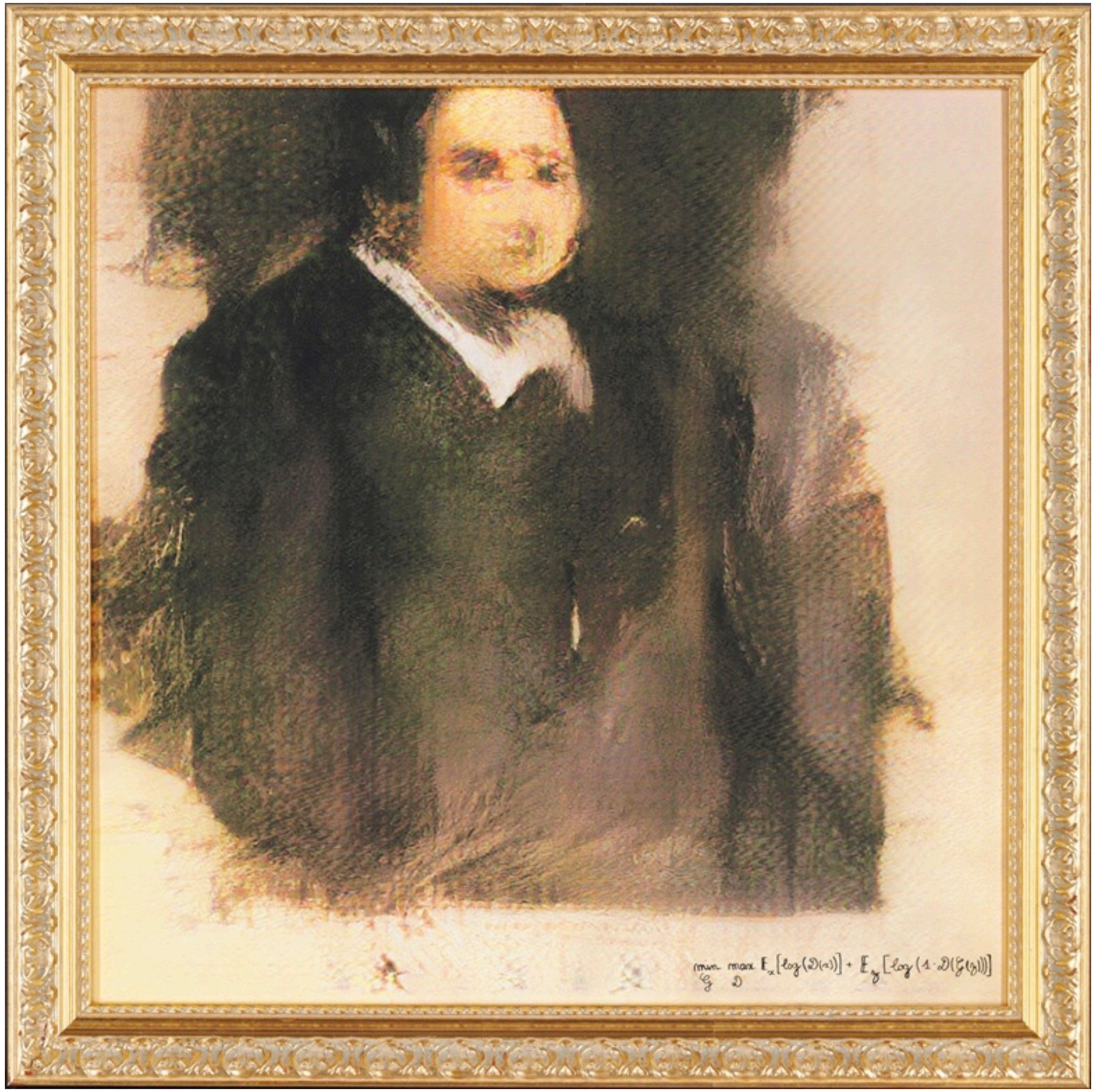
Deep learning is more than function fitting





"What I can not create, I do not understand"

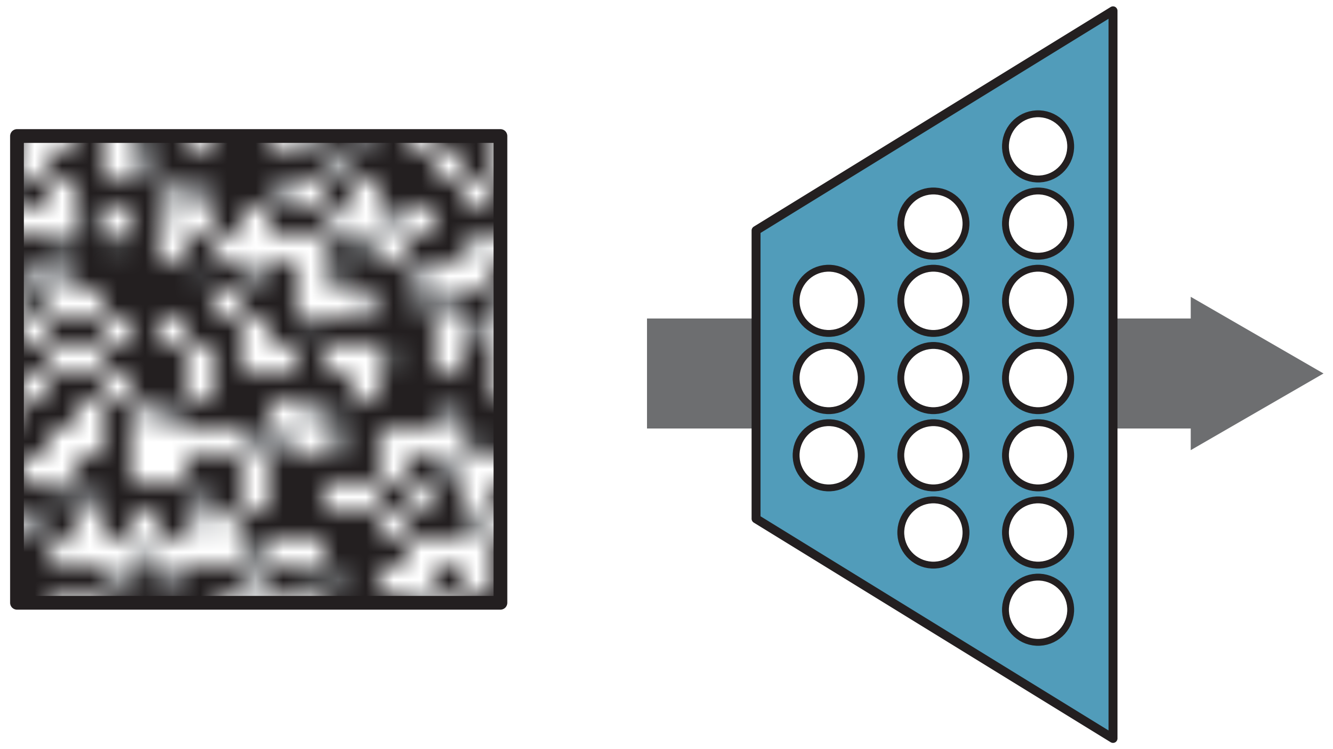
Generated Arts



$432,500
25 October 2018
Christie's New York

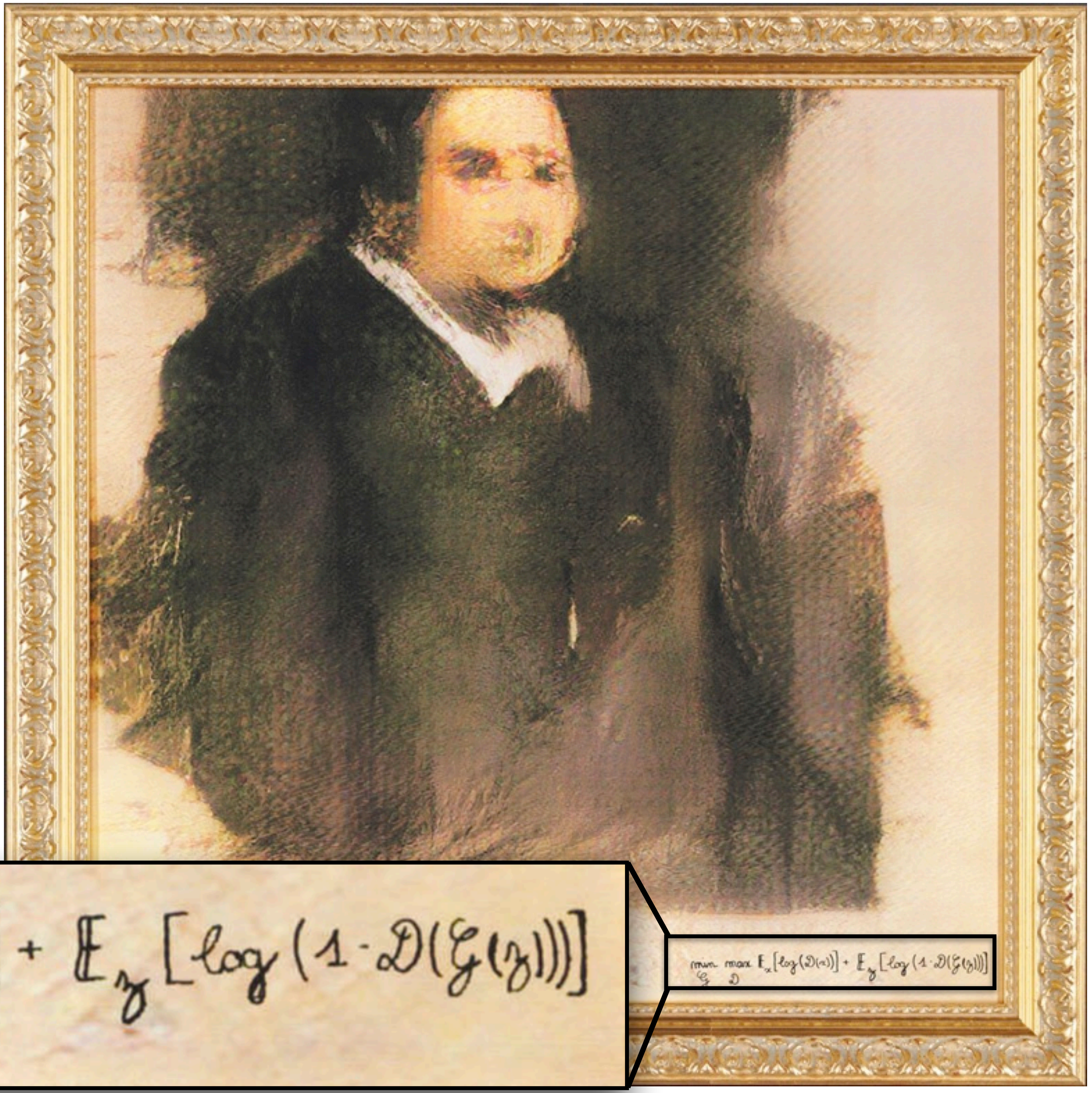
https://www.christies.com/Features/A-collaboration-between-two-artists-one-human-one-a-machine-9332-1.aspx

Generated Arts



Gaussian Noise

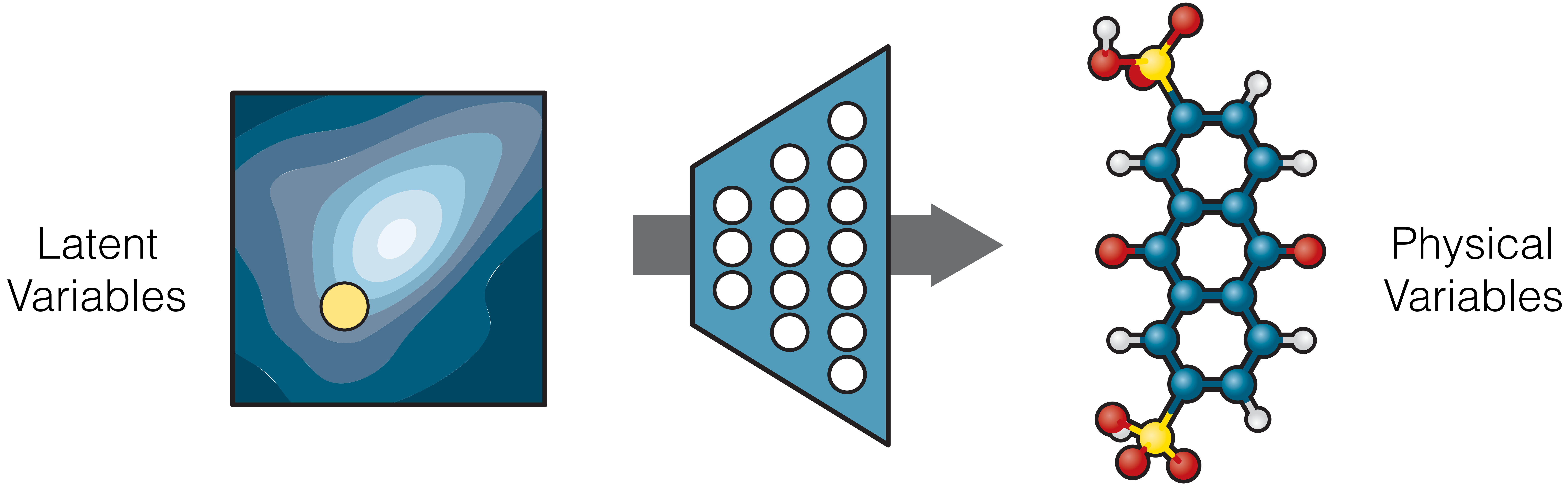
Generative Network



$432,500
25 October 2018
Christie’s New York

https://www.christies.com/Features/A-collaboration-between-two-artists-one-human-one-a-machine-9332-1.aspx

Generate Molecules



Simple Distributions

Generate

Inference

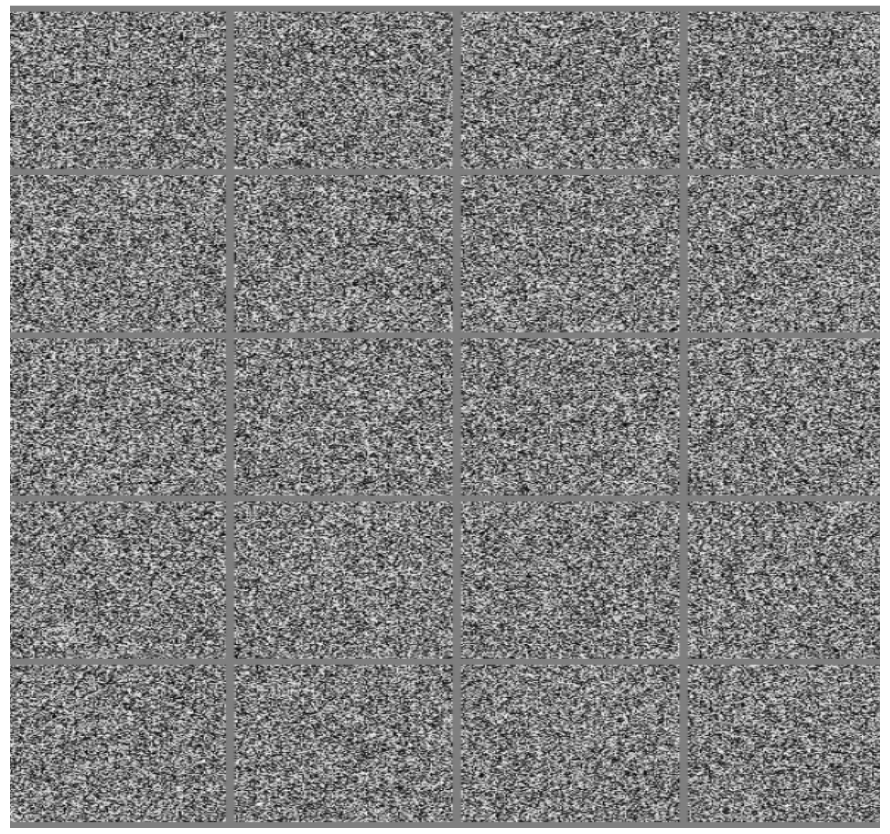
Complex Distribution

Sanchez-Lengeling & Aspuru-Guzik, Science 2018

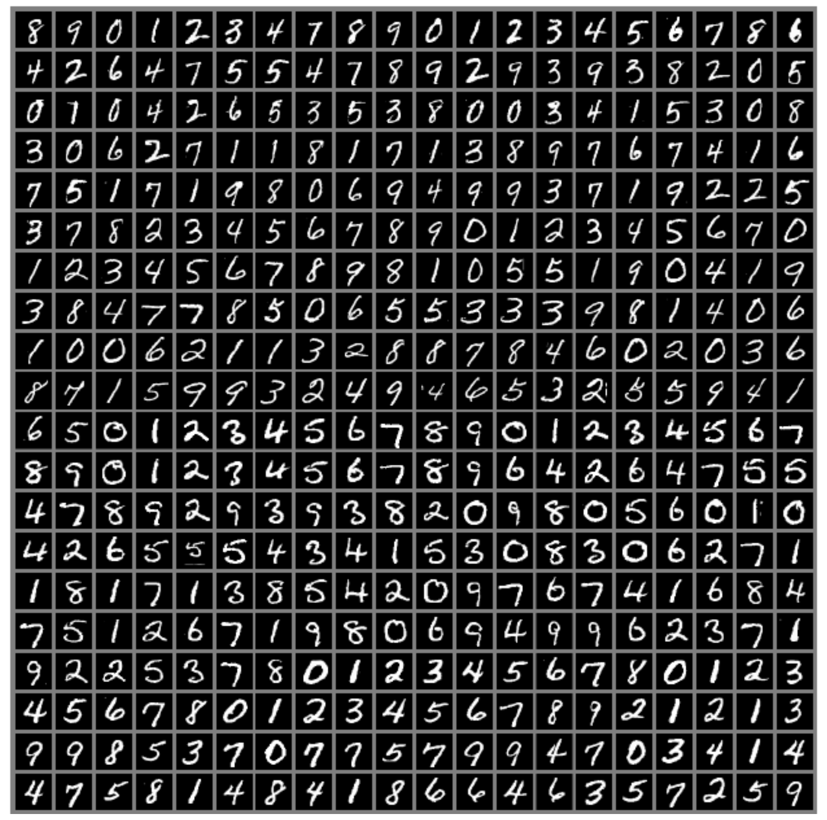
Probabilistic Generative Modeling

$$p(\boldsymbol{x})$$

How to express, learn, and sample from a high-dimensional probability distribution ?



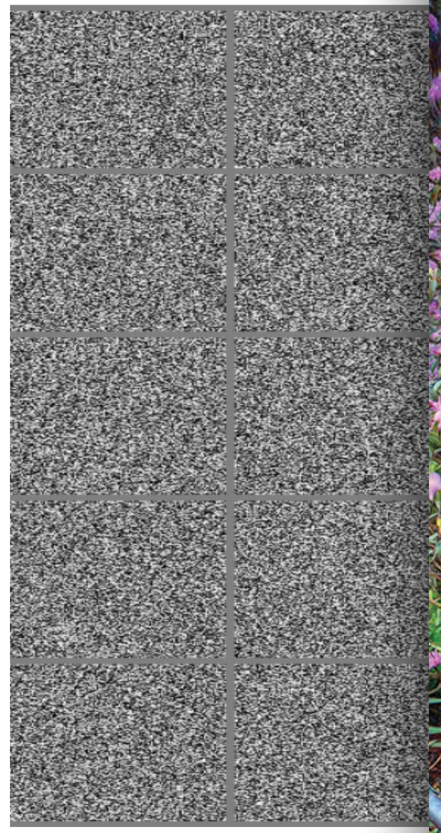
“random” images

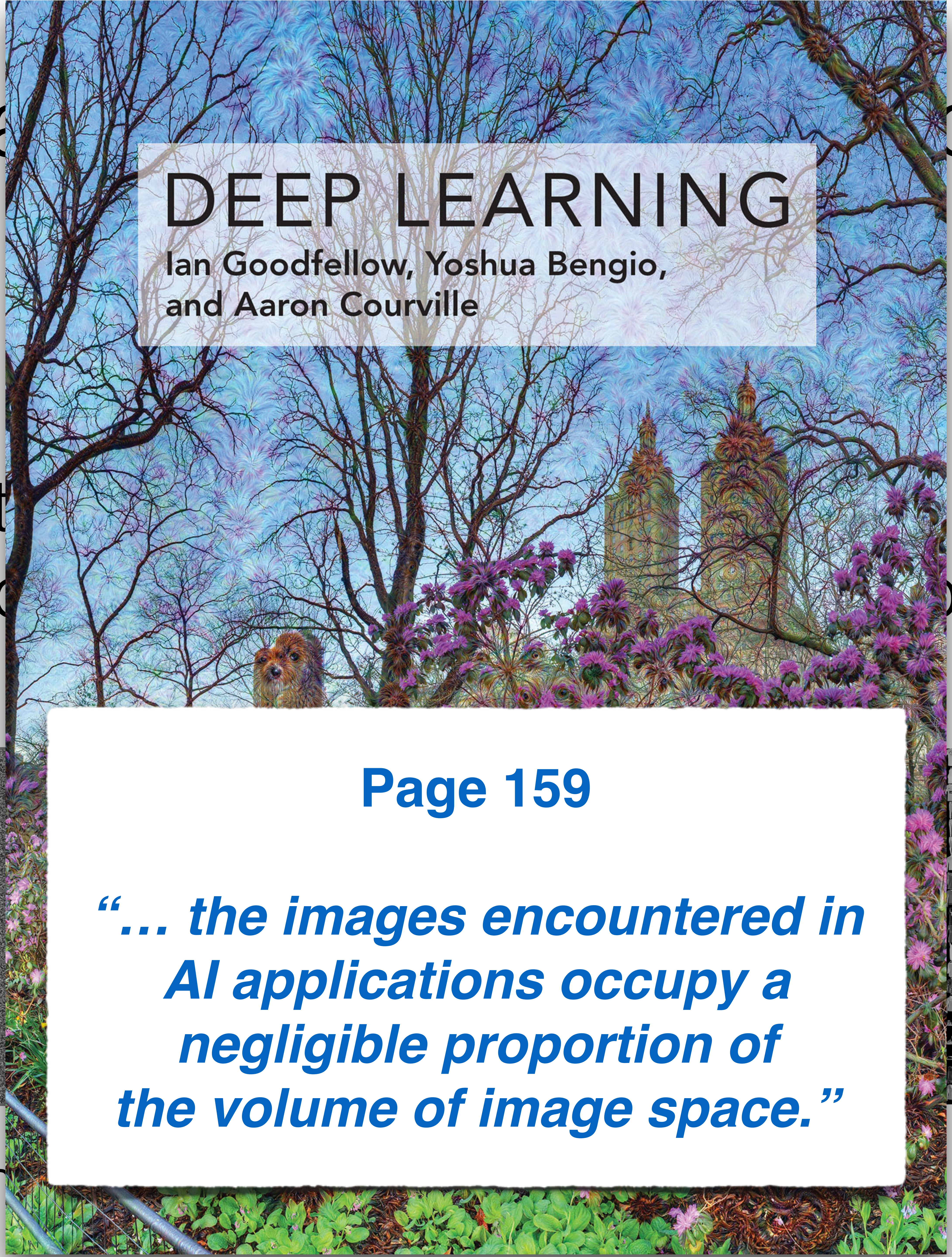


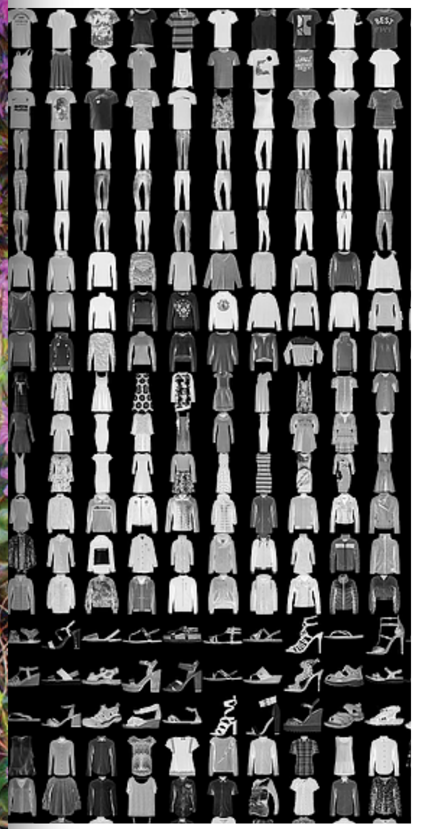
“natural” images

Proba odeling

How t from a
high-(ution ?





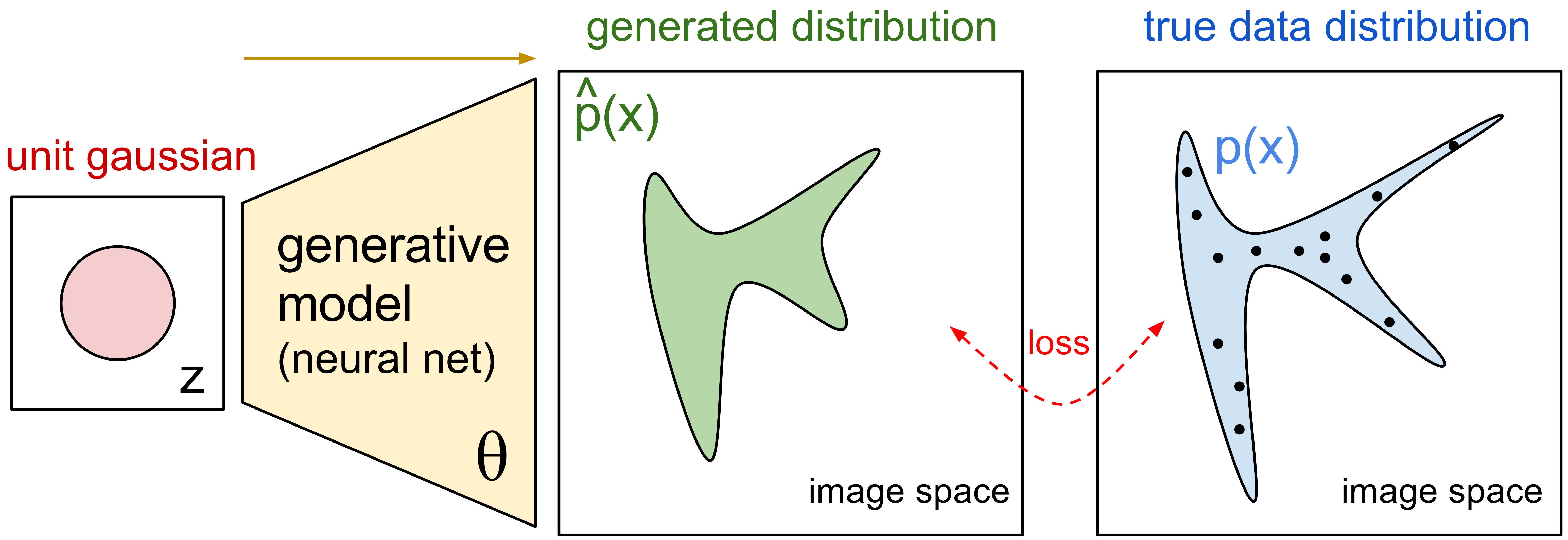


“random

Probabilistic Generative Modeling

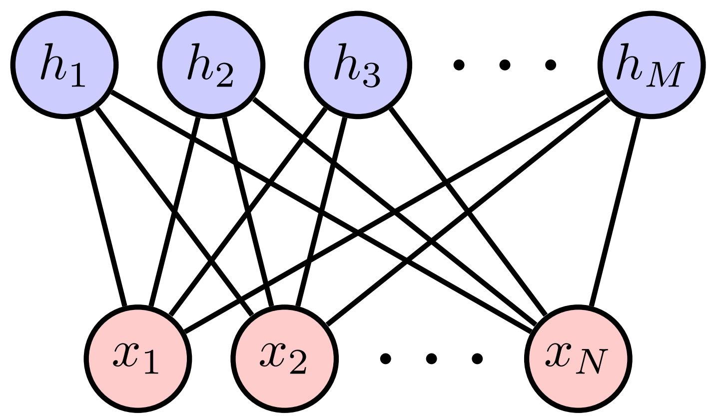
$$p(\boldsymbol{x})$$

How to express, learn, and sample from a high-dimensional probability distribution ?



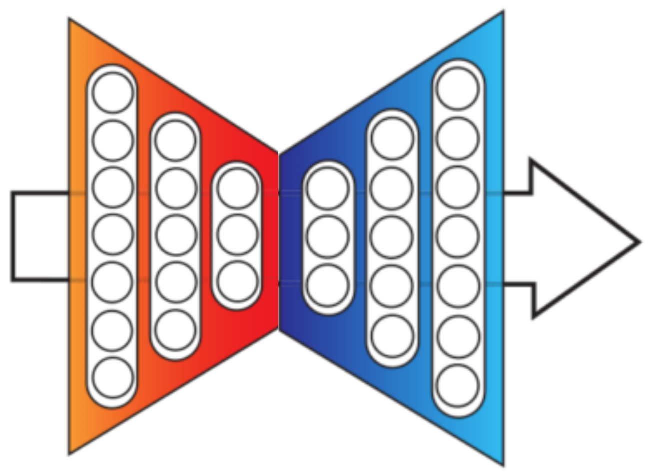
https://blog.openai.com/generative-models/

Timeline of Generative Models



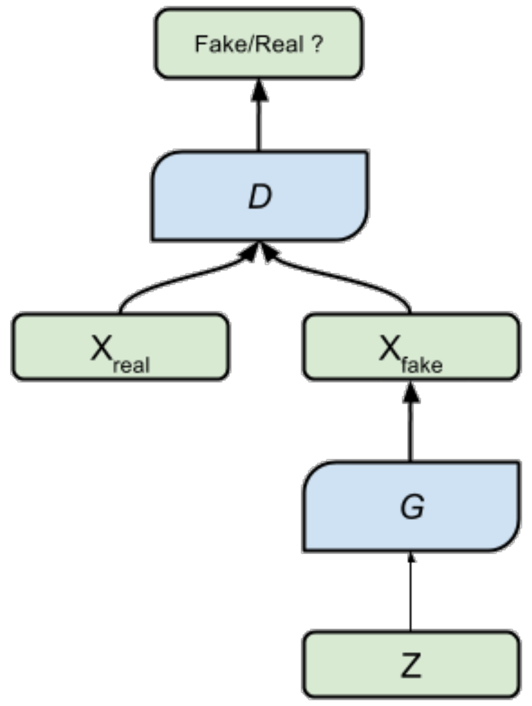
Boltzmann Machines

1980s



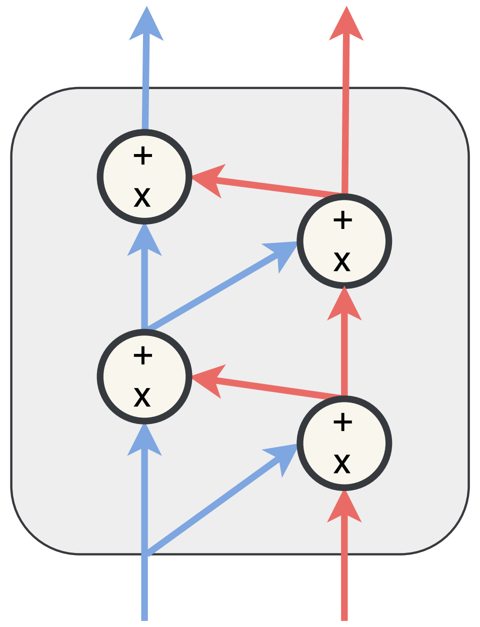
Variational Autoendoer

2013



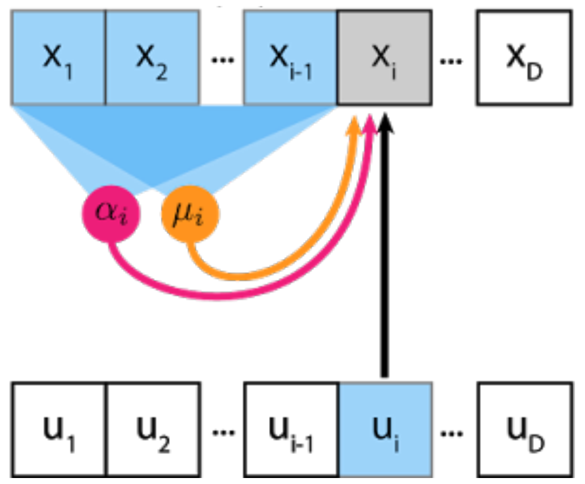
Adversarial Network

2014



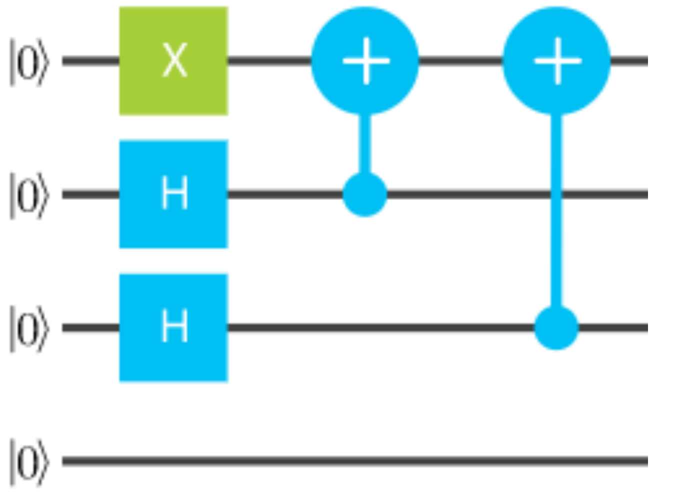
Normalizing Flows

2015



Autoregressive Flows

2016



Born Machines

2017

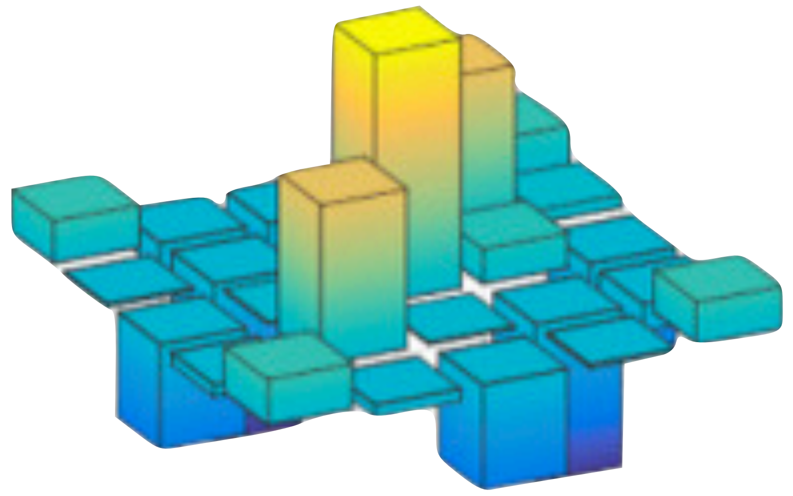
❶ **Leverage the power of modern generative models for physics**

❷ **Statistical, quantum, and fluid mechanics inspired generative models**

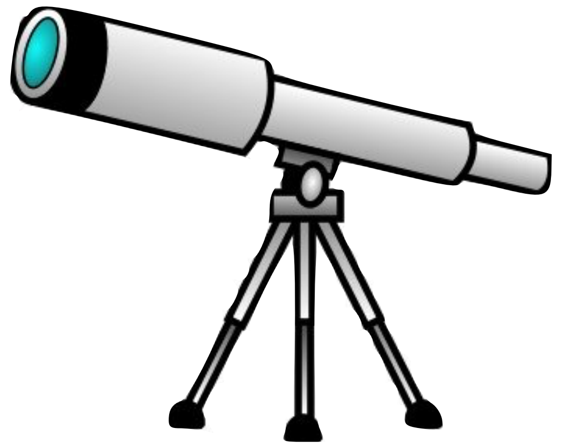
Application of generative models

Ψ

Variational ansatz



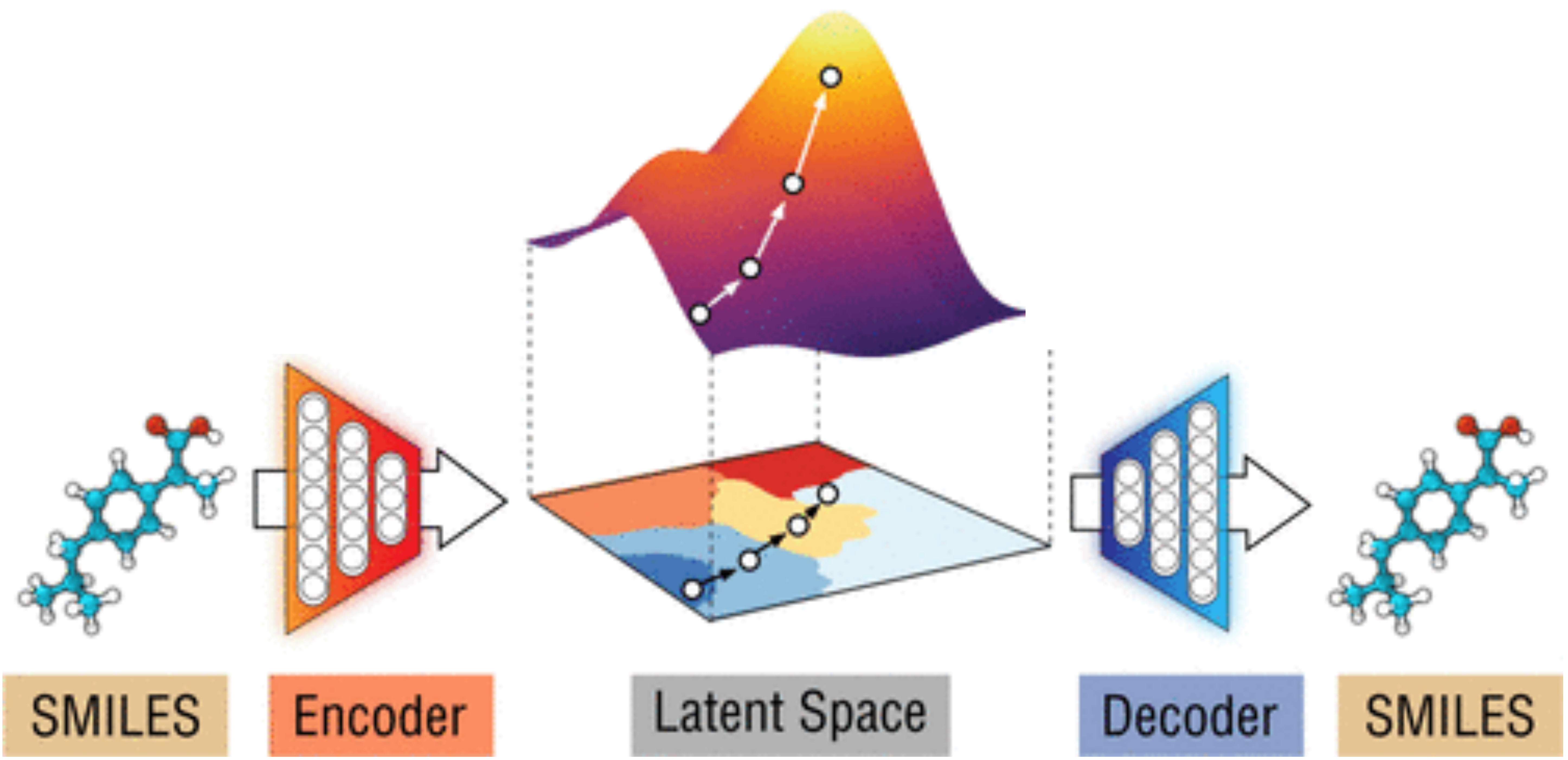
Quantum tomography



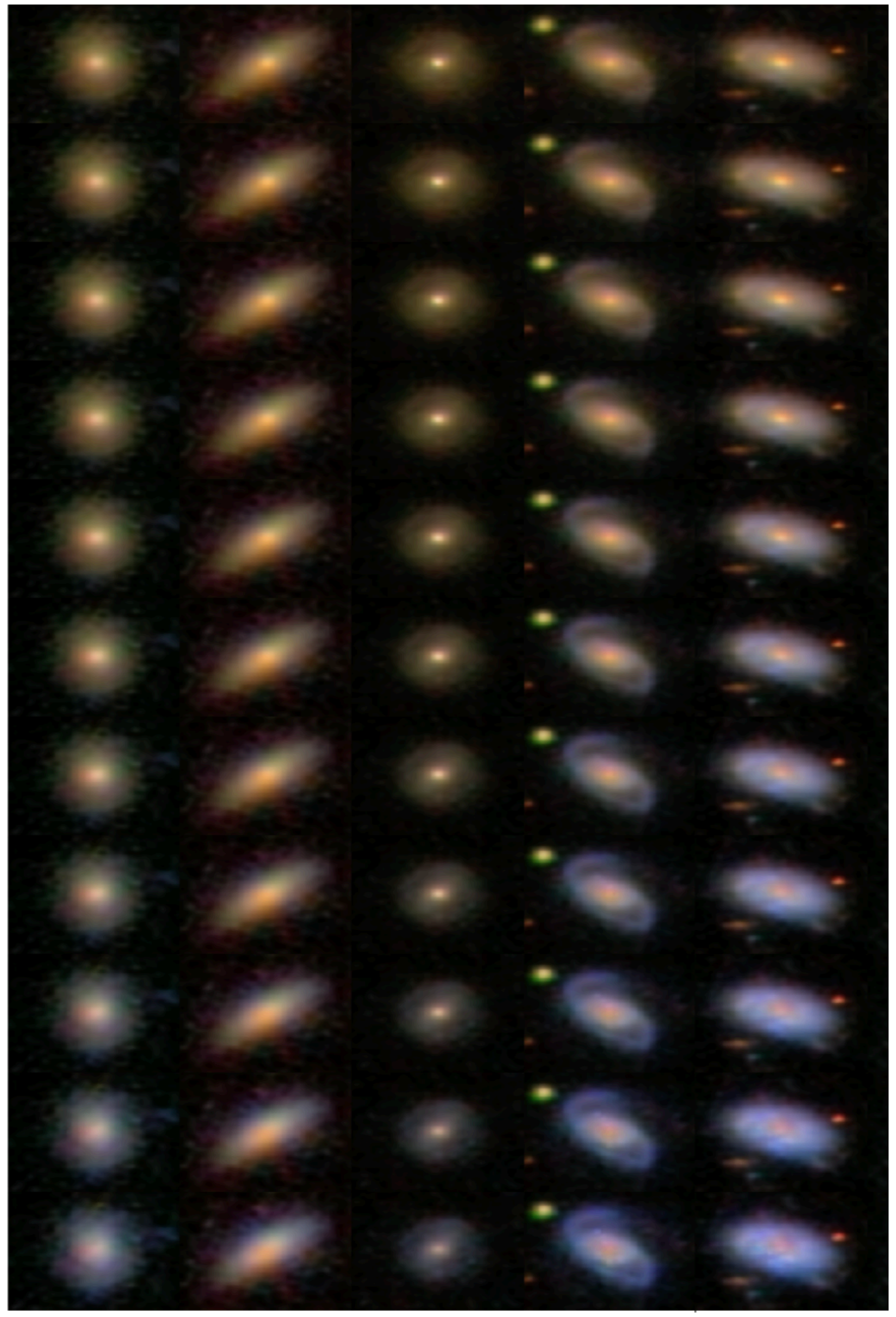
Renormalization group

Monte Carlo update

Application of generative models



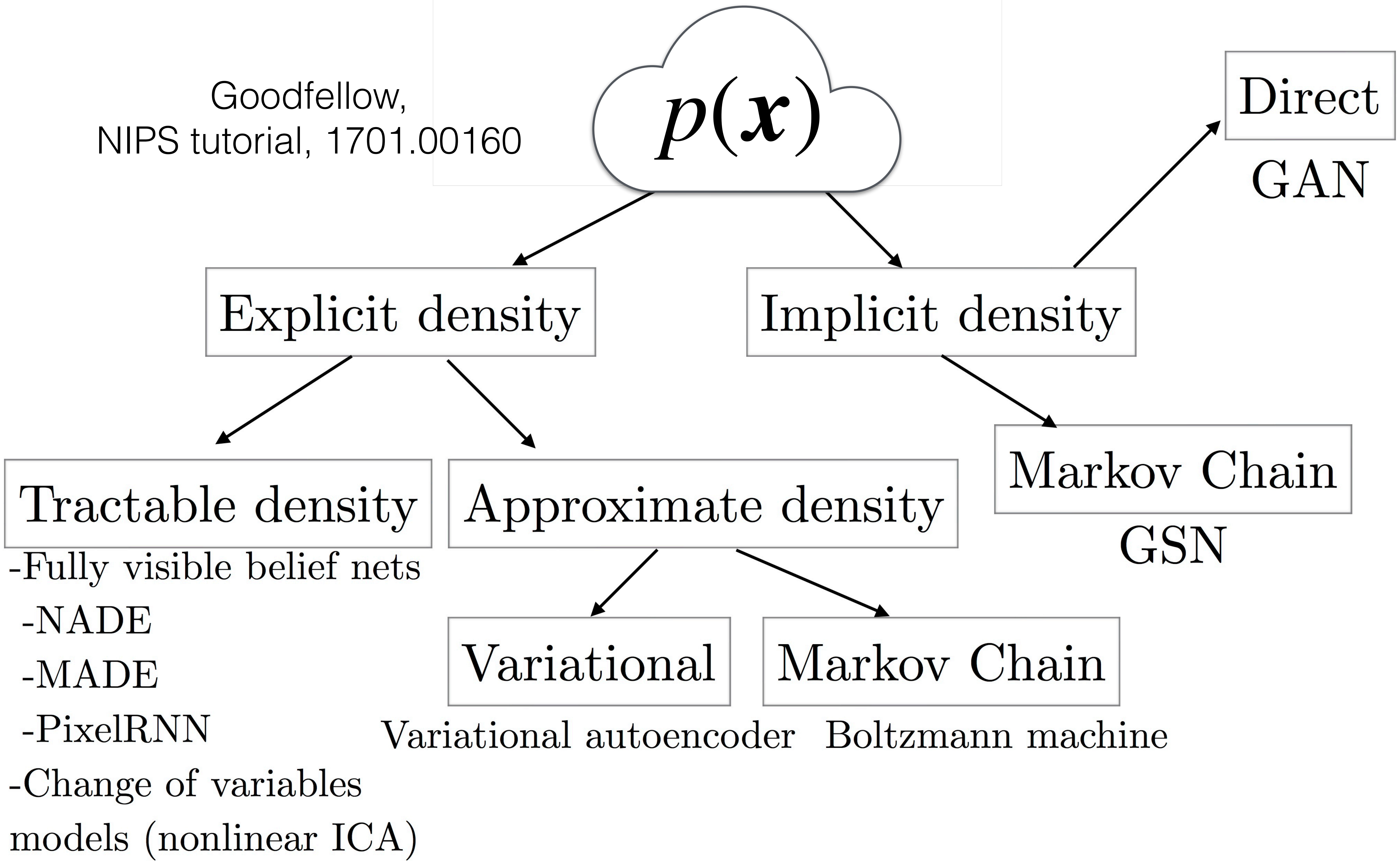
Automatic chemical design,
Gomez-Bombarelli et al, 1610.02415



Galaxy evolution
Schawinski et al, unpublished

Physics genes of generative models

Goodfellow,
NIPS tutorial, 1701.00160

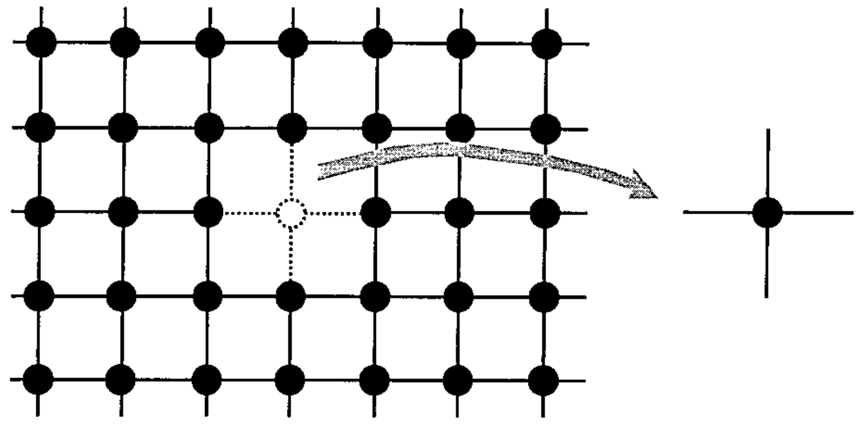


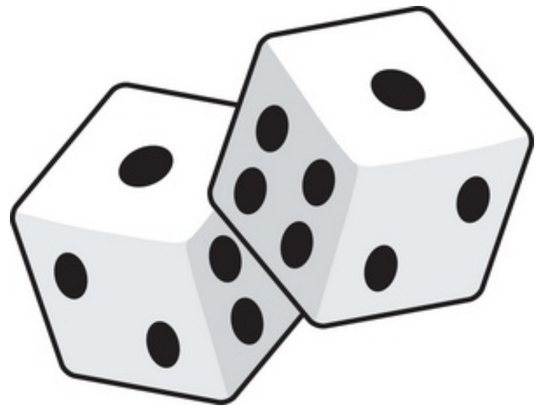
Physics genes of generative models

Goodfellow,
NIPS tutorial, 1701.00160

$p(\boldsymbol{x})$

- Explicit density
  - Tractable density
    - -Fully visible belief nets
      - -NADE
      - -MADE
      - -PixelRNN
    - -Change of variables models (nonlinear ICA)
  - Approximate density
    - Variational
      - Variational autoencoder
    - Markov Chain
      - Boltzmann machine
- Implicit density
  - Direct
    - GAN
  - Markov Chain
    - GSN



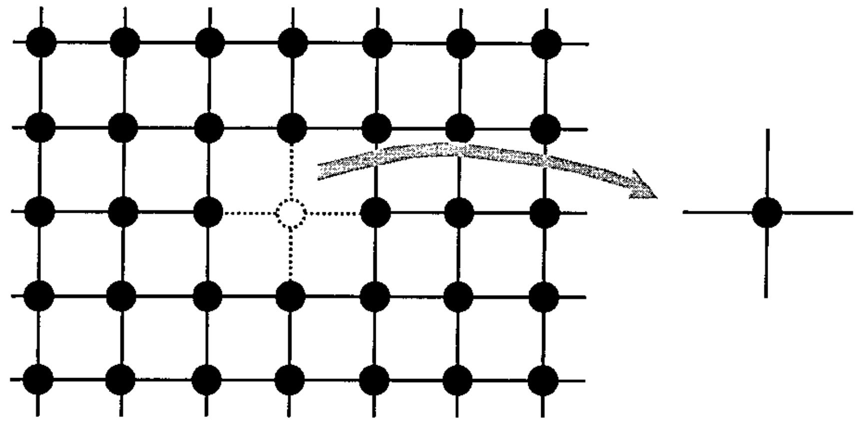


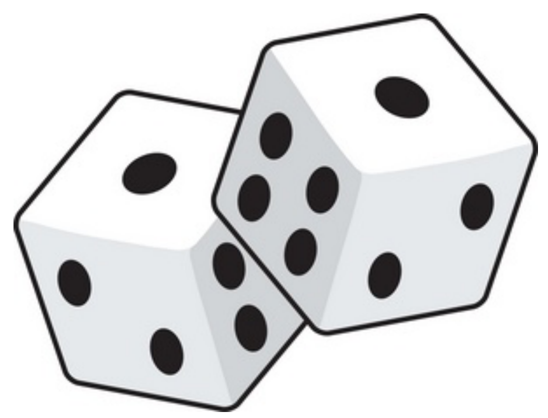
Physics genes of generative models

Goodfellow,
NIPS tutorial, 1701.00160

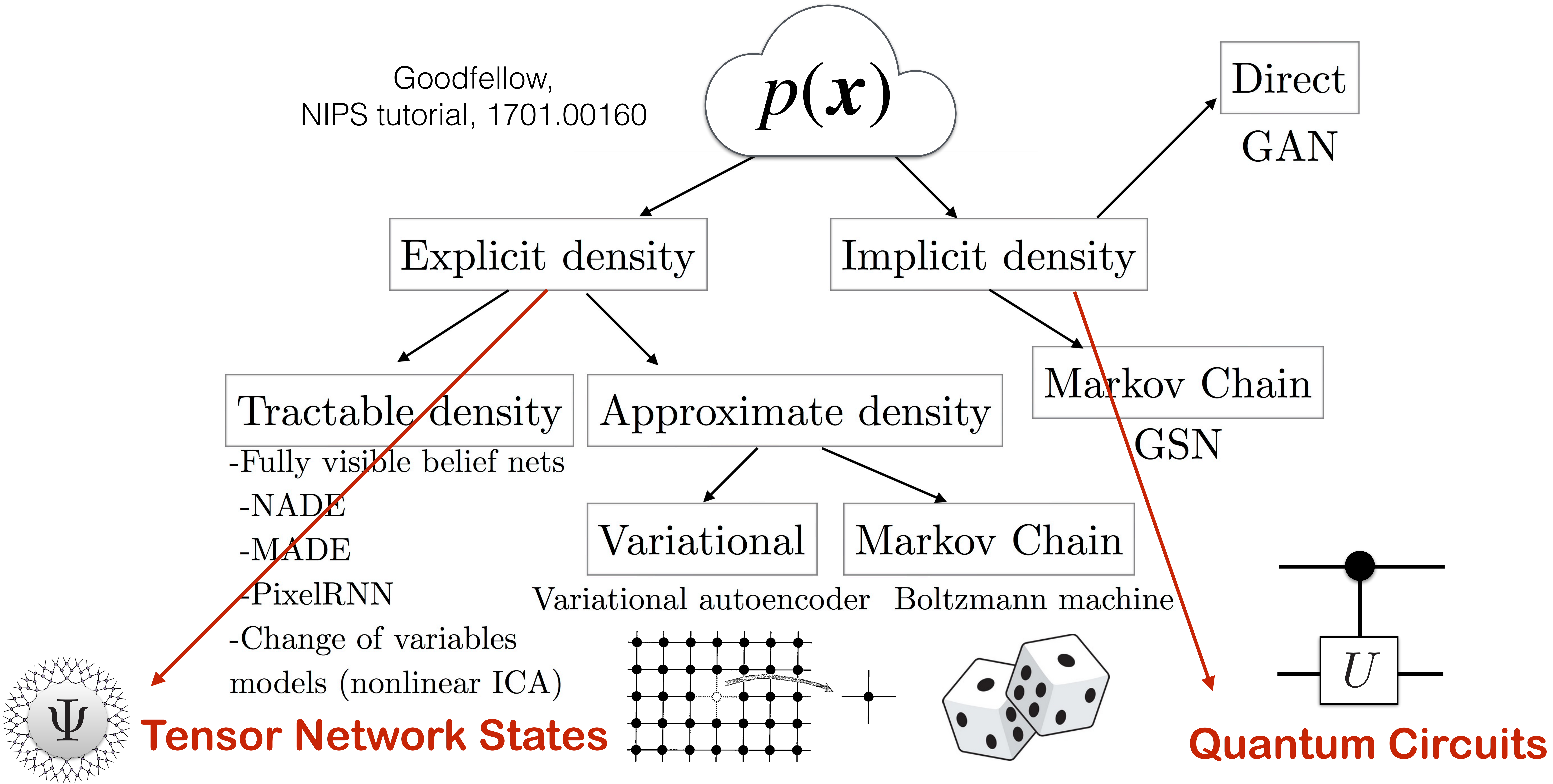
$p(\boldsymbol{x})$

- Explicit density
  - Tractable density
    - -Fully visible belief nets
      - -NADE
      - -MADE
      - -PixelRNN
    - -Change of variables models (nonlinear ICA)
    - **Tensor Network States**
  - Approximate density
    - Variational
      - Variational autoencoder
    - Markov Chain
      - Boltzmann machine
- Implicit density
  - Direct
    - GAN
  - Markov Chain
    - GSN





Physics genes of generative models



DL as a fluid control problem

$$\frac{p(\boldsymbol{z})}{q(\nabla u(\boldsymbol{z}))} = \det\left(\frac{\partial^2 u}{\partial z_i \partial z_j}\right)$$

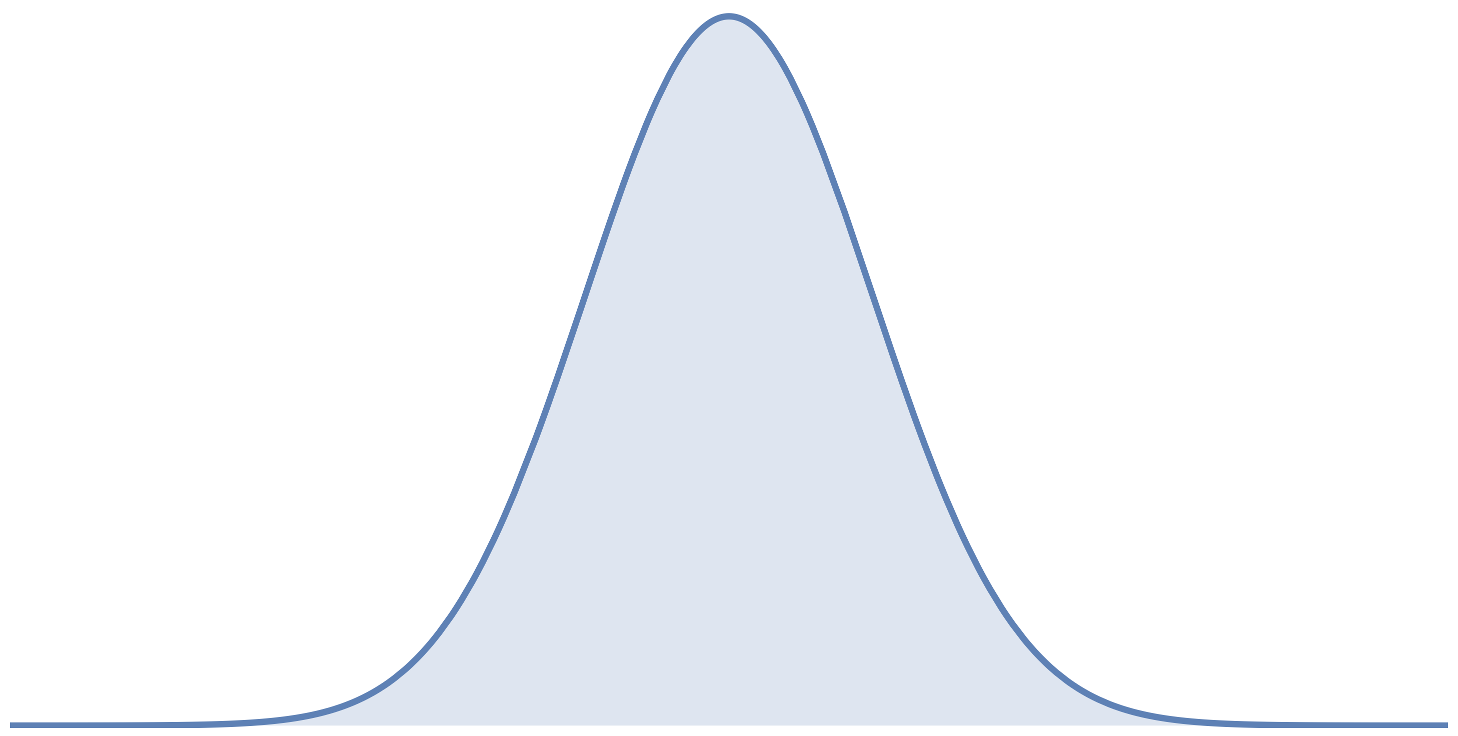
Monge-Ampère equation
in optimal transport theory

Continuous-time limit

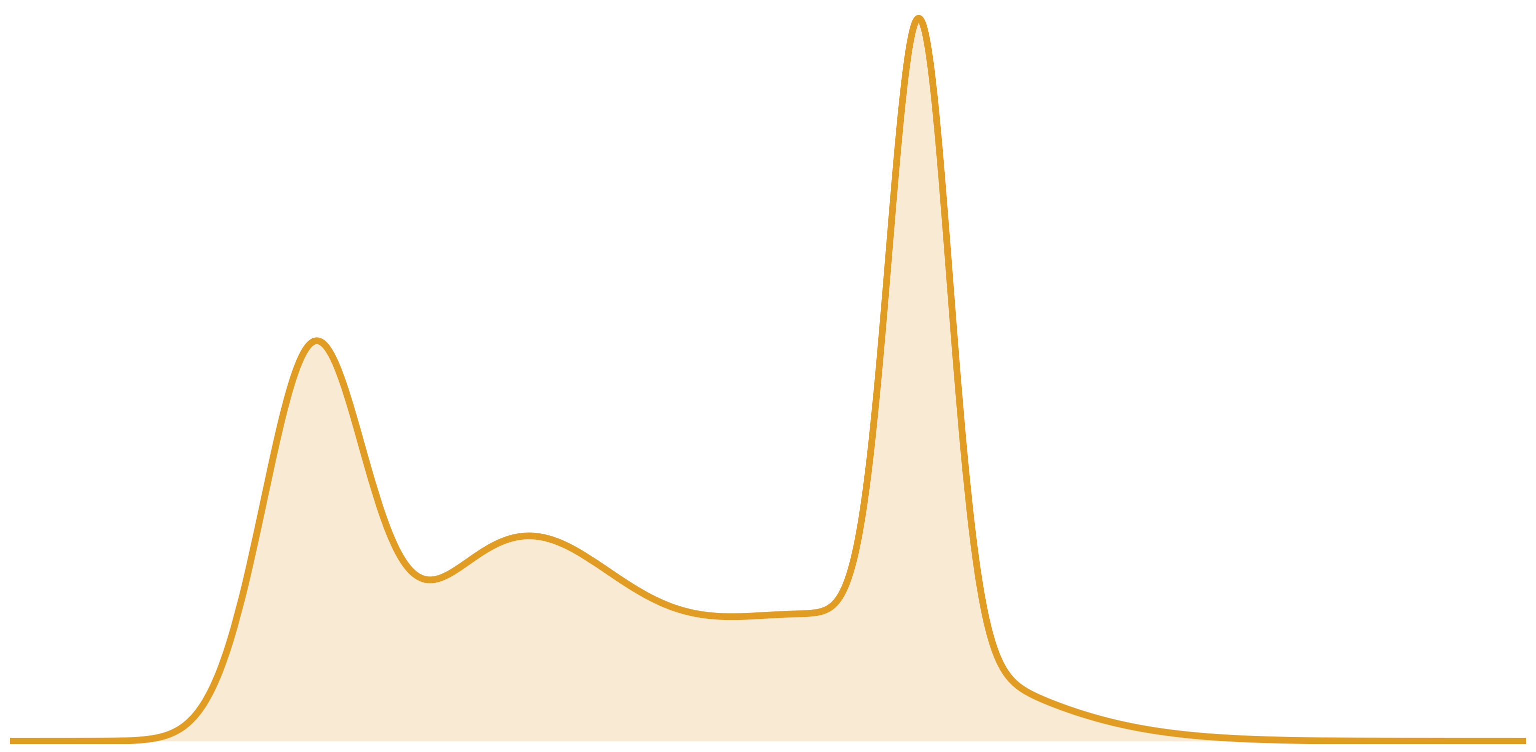
$$\xrightarrow[\substack{u(\boldsymbol{z}) = |\boldsymbol{z}|^2/2 + \epsilon\varphi(\boldsymbol{z}) \\ \epsilon \to 0}]{}$$

$$\frac{\partial p(\boldsymbol{x}, t)}{\partial t} + \nabla \cdot \left[p(\boldsymbol{x}, t) \nabla \varphi\right] = 0$$

Continuity equation of
compressible fluids



Simple density



Complex density

Zhang, E, LW, 1809.10188
c.f. Neural ODE, 1806.07366

DL as a fluid control problem

$$\frac{p(\mathbf{z})}{q(\nabla u(\mathbf{z}))} = \det\left(\frac{\partial^2 u}{\partial z_i \partial z_j}\right)$$

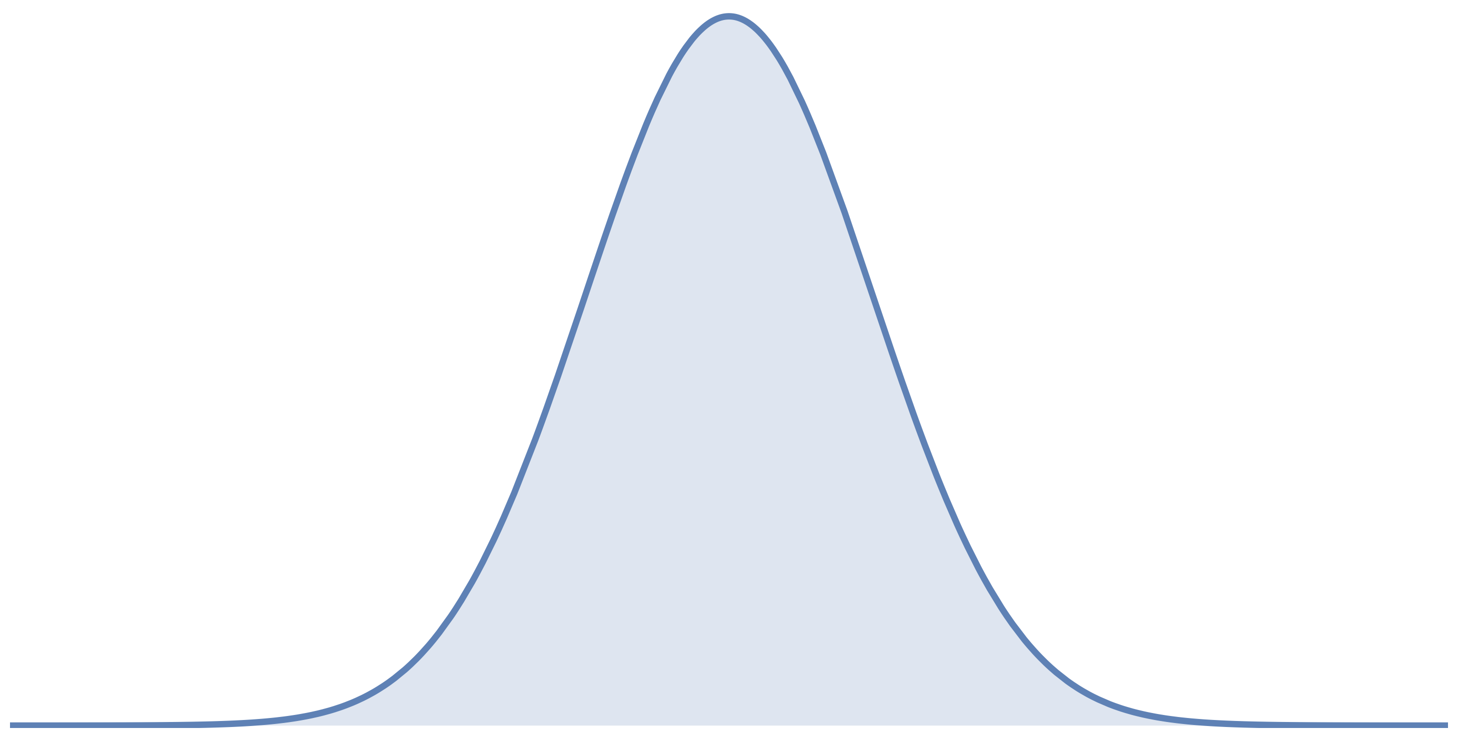
Monge-Ampère equation in optimal transport theory

Continuous-time limit

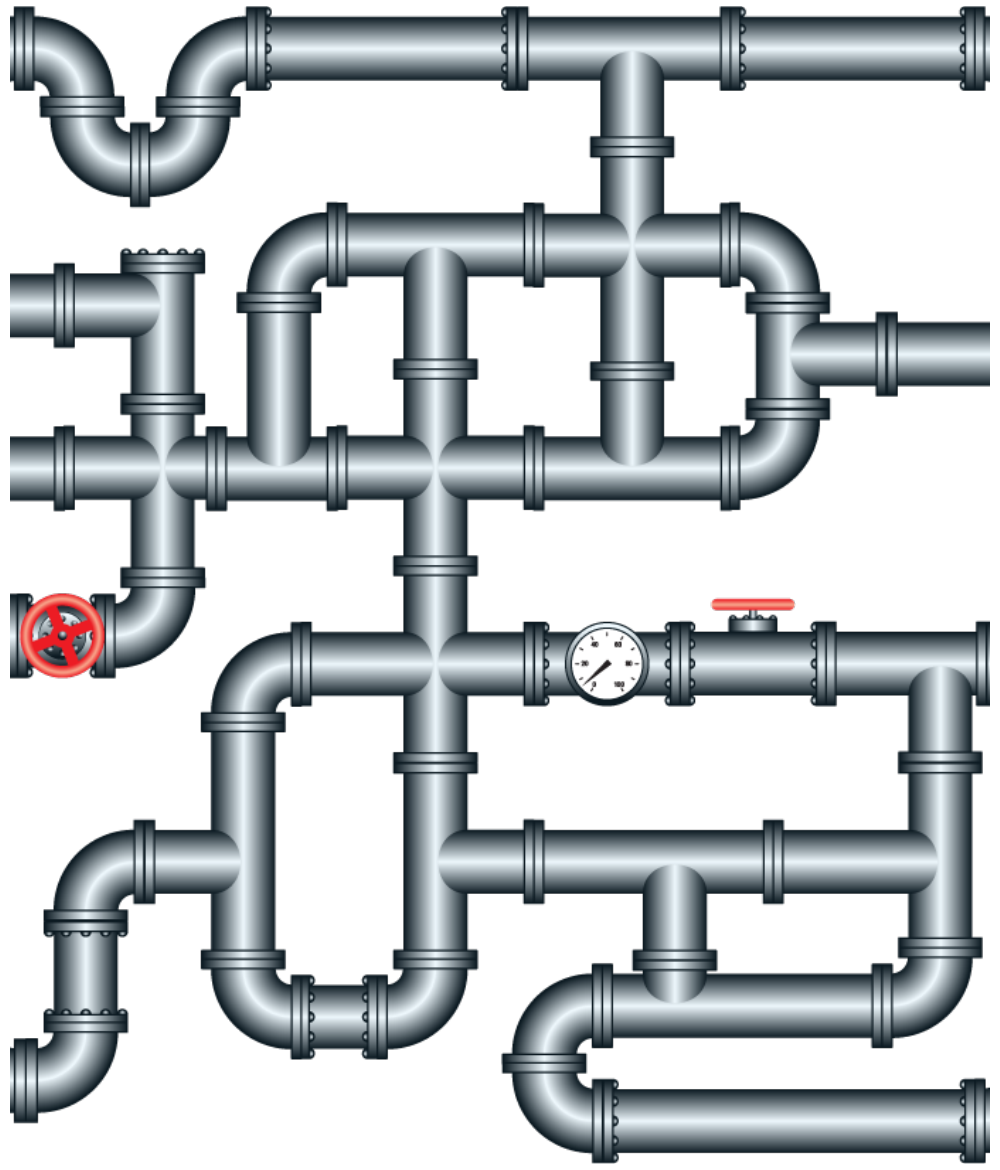
$$u(\mathbf{z}) = |\mathbf{z}|^2/2 + \epsilon\varphi(\mathbf{z})$$
$$\epsilon \to 0$$

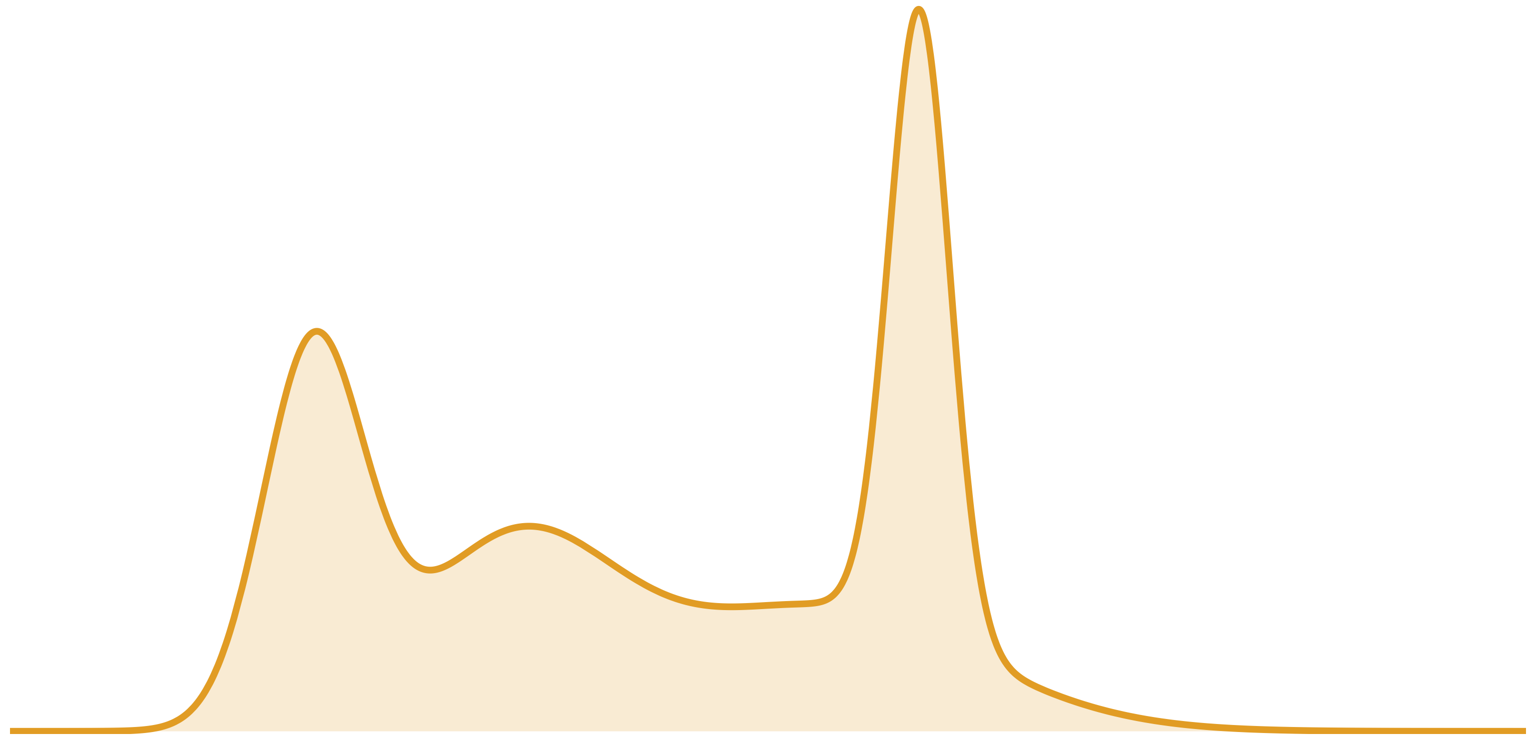
$$\frac{\partial p(\boldsymbol{x}, t)}{\partial t} + \nabla \cdot \left[p(\boldsymbol{x}, t)\nabla\varphi\right] = 0$$

Continuity equation of compressible fluids



Simple density





Complex density

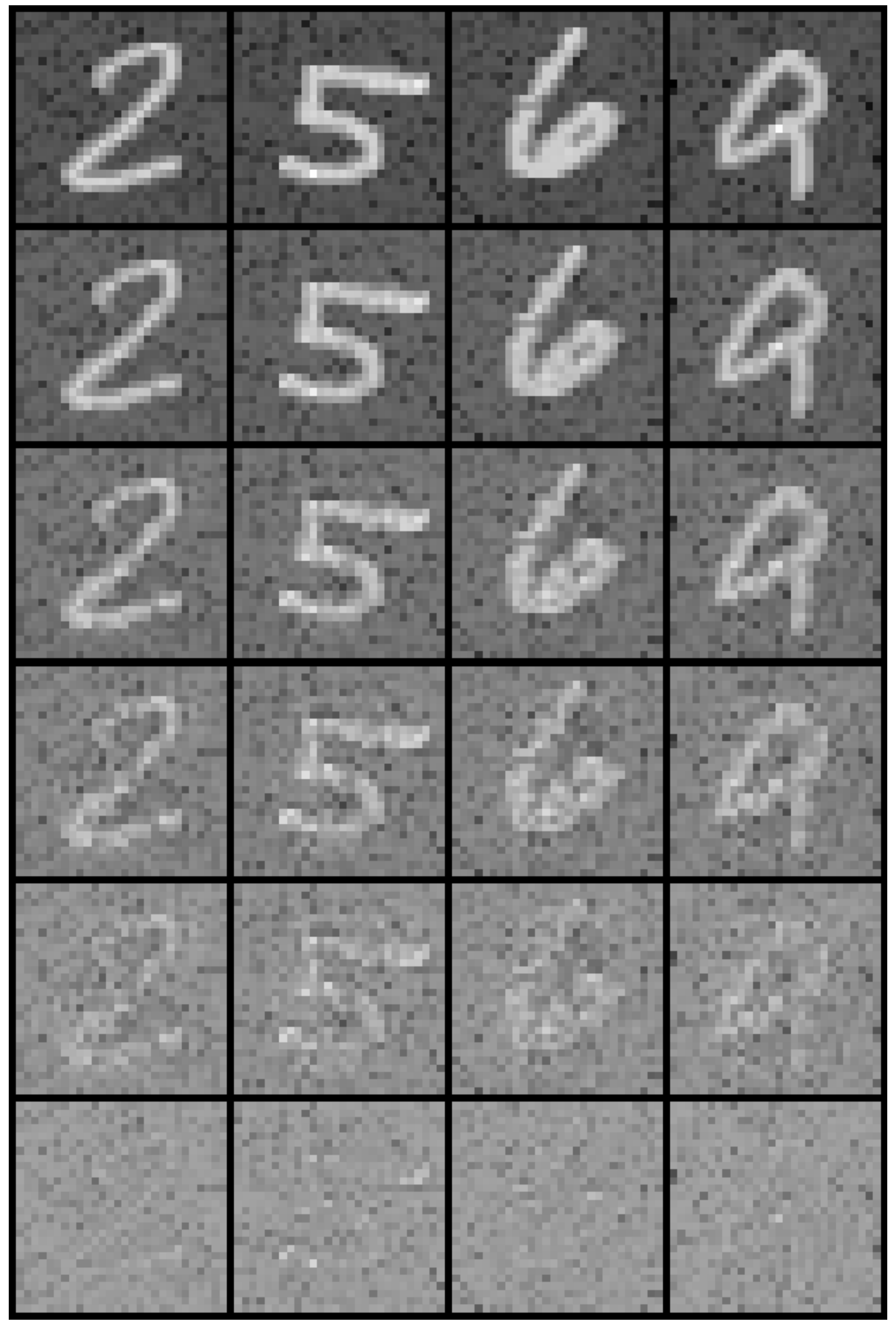
Zhang, E, LW, 1809.10188

c.f. Neural ODE, 1806.07366

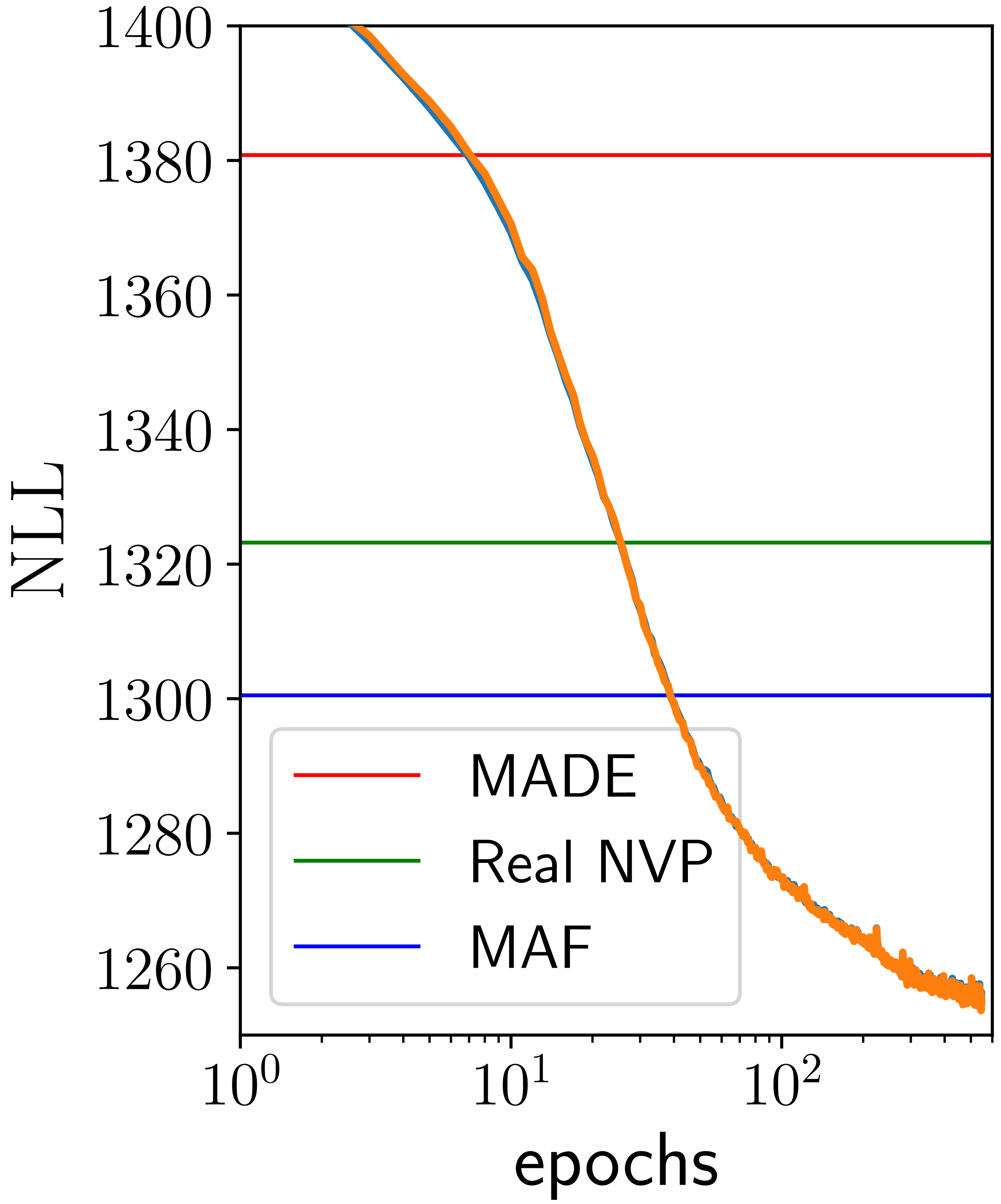
Density estimation of hand-written digits

A standard benchmark for generative models, lower is better

Data space



Latent space



'15

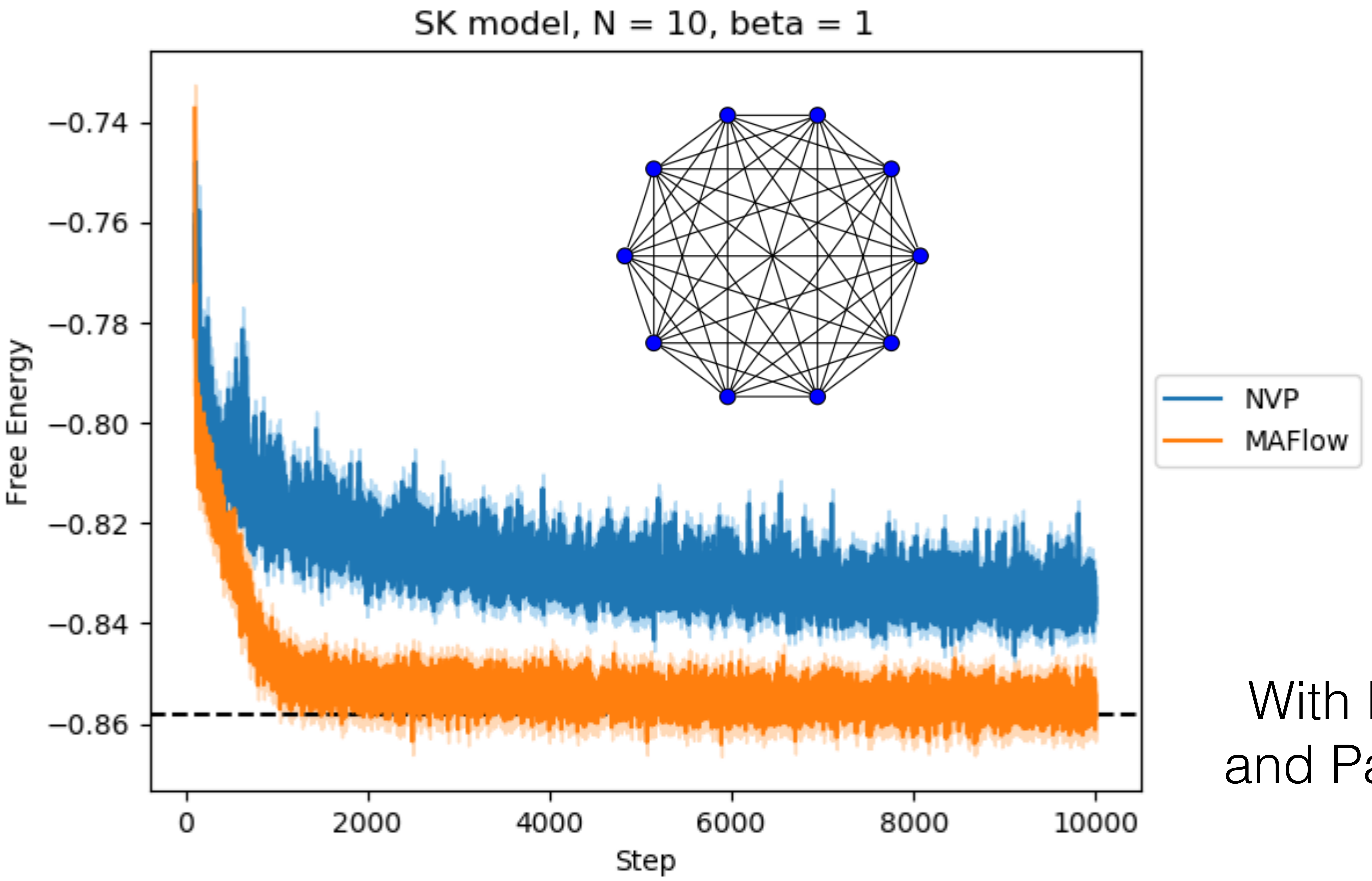
'16

'17

'18 Our Result

State-of-the-art performance in unstructured density estimation

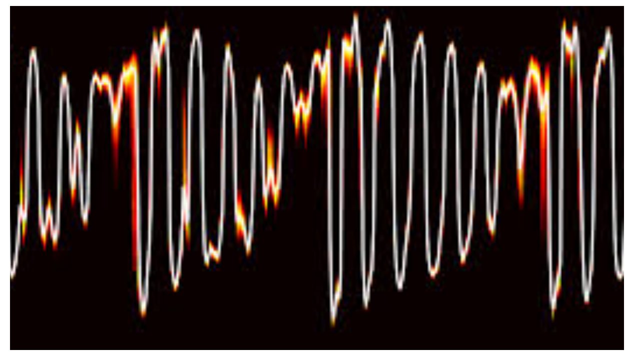
Variational study of spin glasses

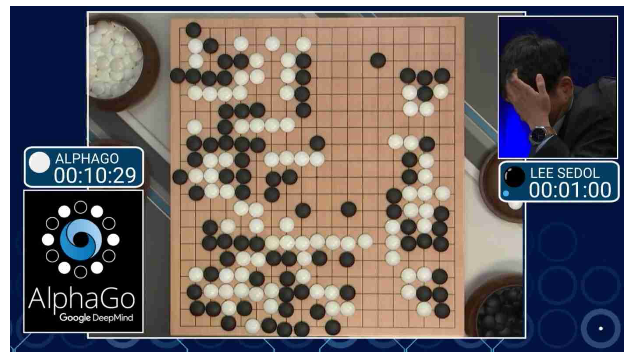


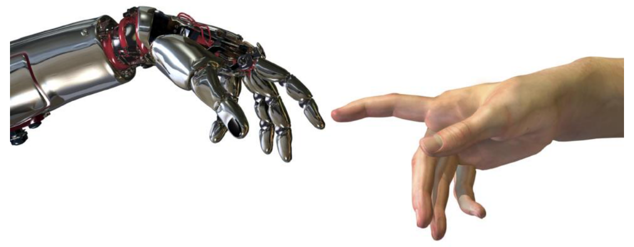
With Dian Wu and Pan Zhang

Better variational energy than previous architectures

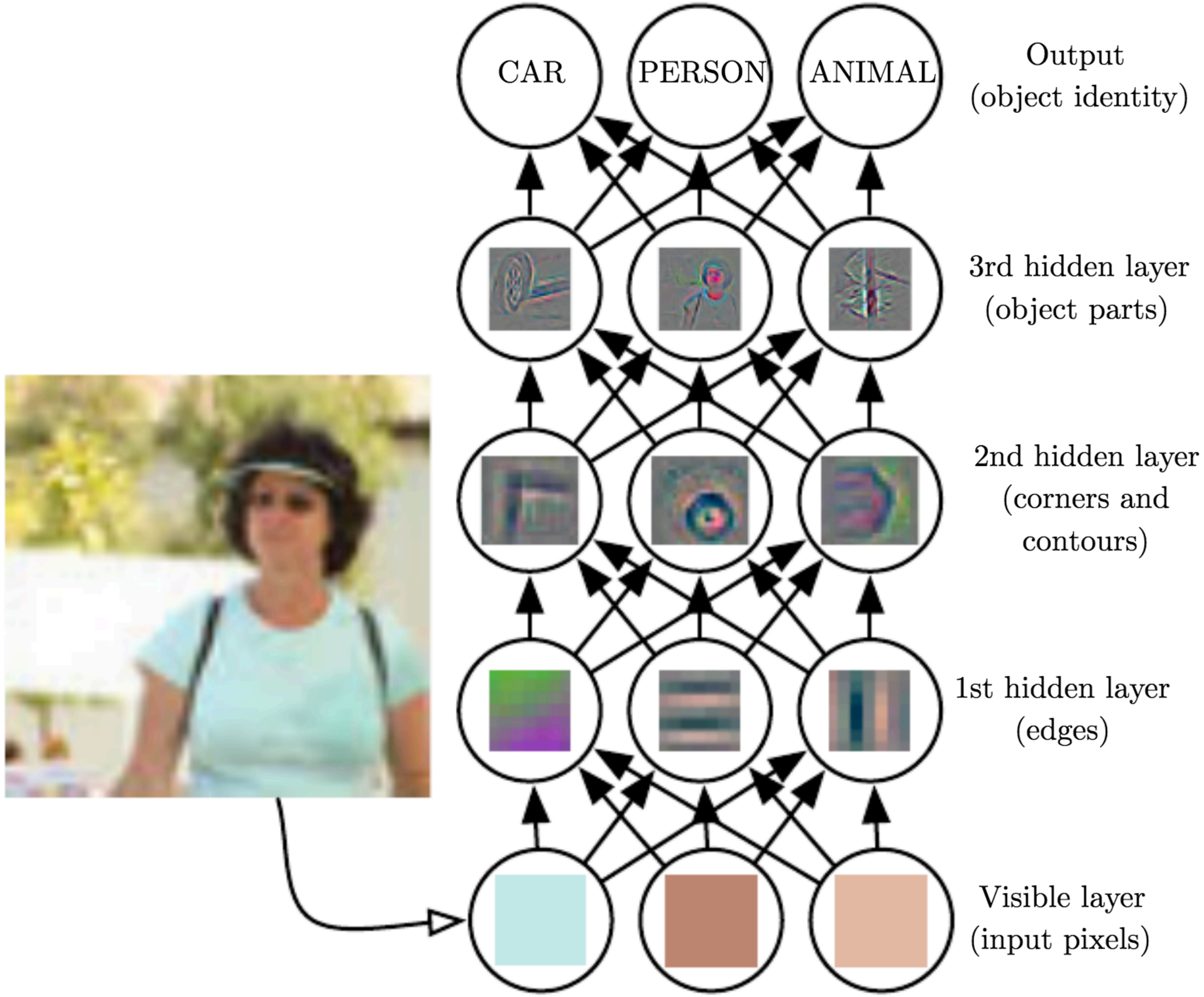
What is the secret behind deep learning?



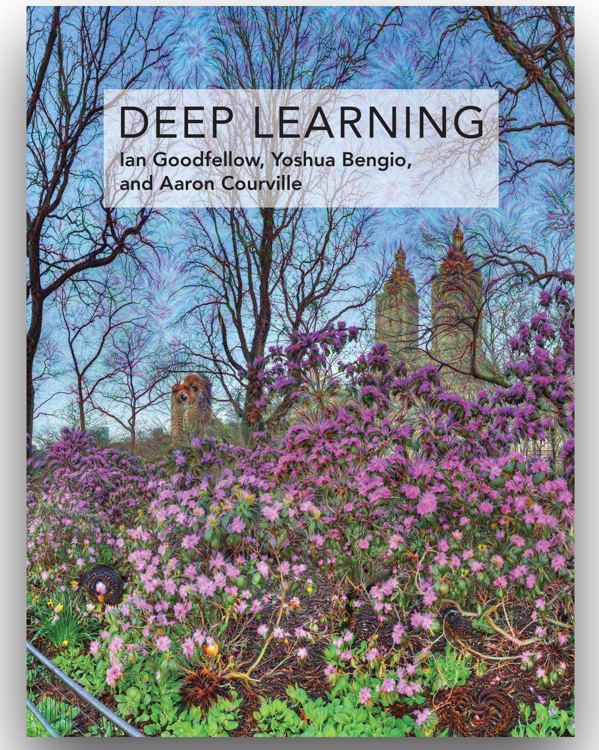




Representation Leanring



Page 6
Figure 1.2



Goodfellow, Bengio, Courville, http://www.deeplearningbook.org/

Magic of learned representations

Neural style transfer

Gatys et al, 1508.06576

Latent space interpolation

Glow 1807.03039

https://blog.openai.com/glow/

Magic of learned representations

Neural style transfer

Gatys et al, 1508.06576

Latent space interpolation

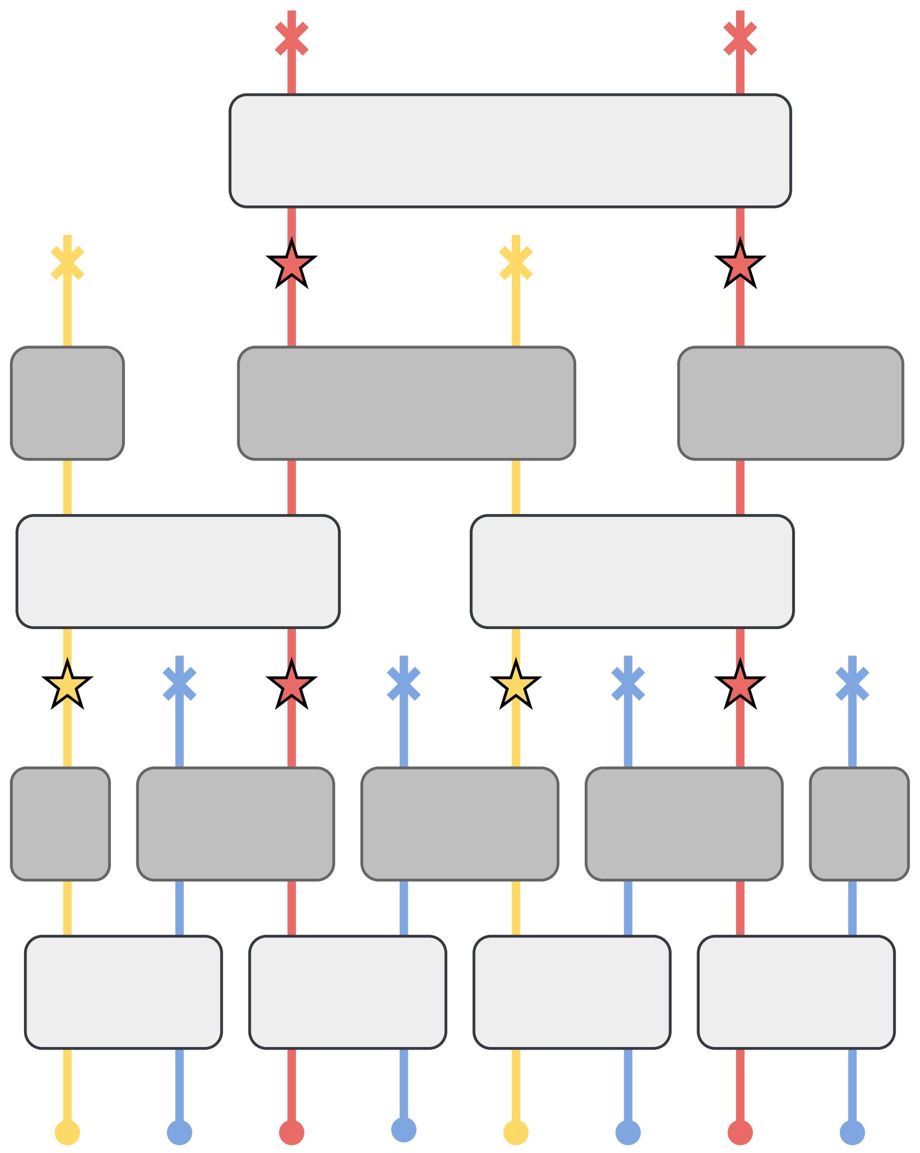
Glow 1807.03039

https://blog.openai.com/glow/

Latent space Hybrid MC

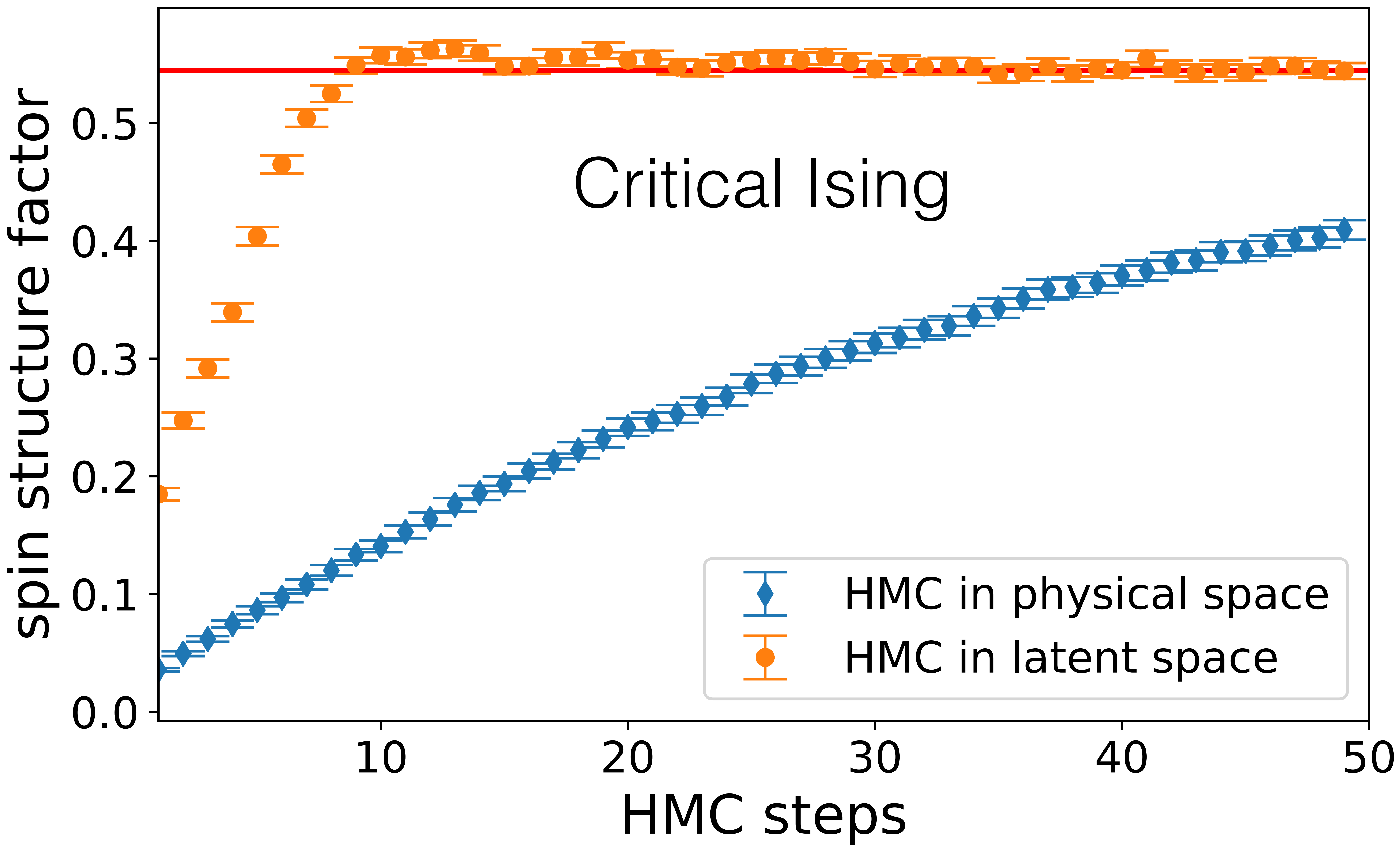
Latent space energy function

$$E_{\text{eff}}(z) = E(g(z)) + \ln q(g(z)) - \ln p(z)$$



Physical energy function $E(\boldsymbol{x})$

NeuralRG, Shuo-Hui Li and LW, 1802.02840

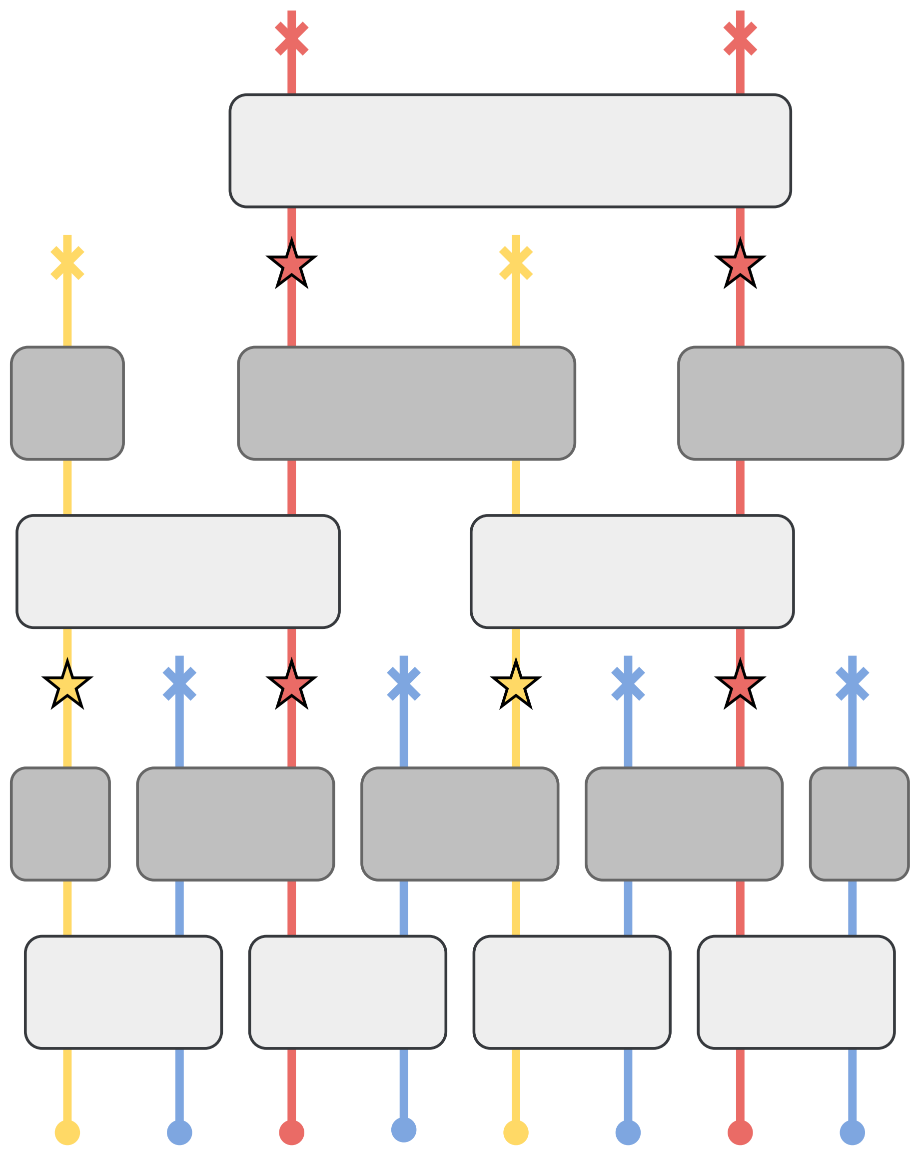


HMC thermalizes faster in the latent space

Latent space Hybrid MC

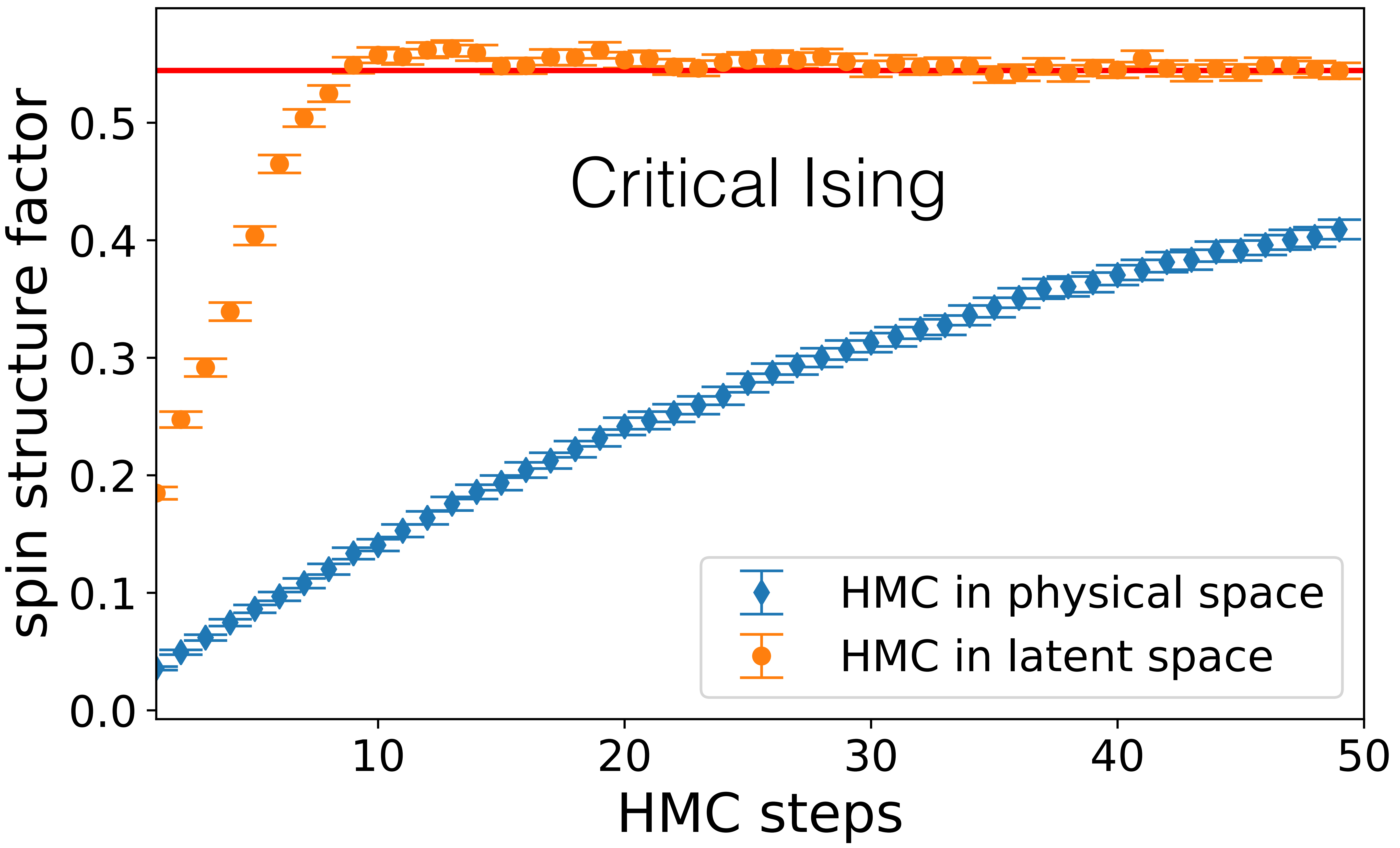
Latent space energy function

$$E_{\text{eff}}(z) = E(g(z)) + \ln q(g(z)) - \ln p(z)$$



Physical energy function $E(\boldsymbol{x})$

NeuralRG, Shuo-Hui Li and LW, 1802.02840

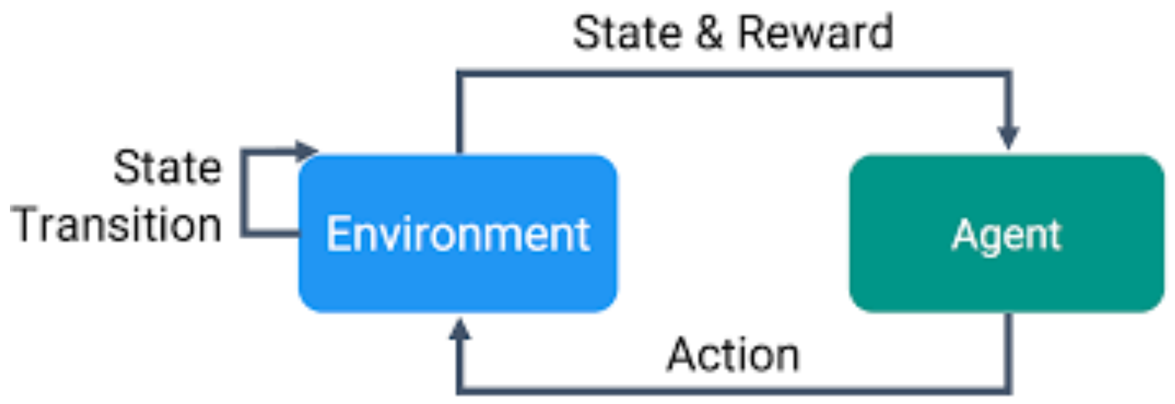


HMC thermalizes faster in the latent space

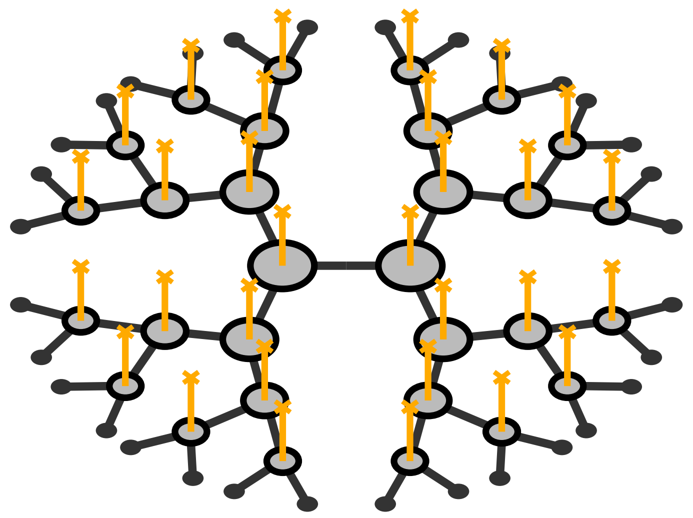
Remarks on accelerated MC

1. Cheap surrogate function for MC weight Neal 96' Jun. S Liu 01' A recommender engine for MC updates when the surrogate is a generative model: Huang, LW, 1610.02746, Liu, Qi, Meng, Fu, 1610.03137

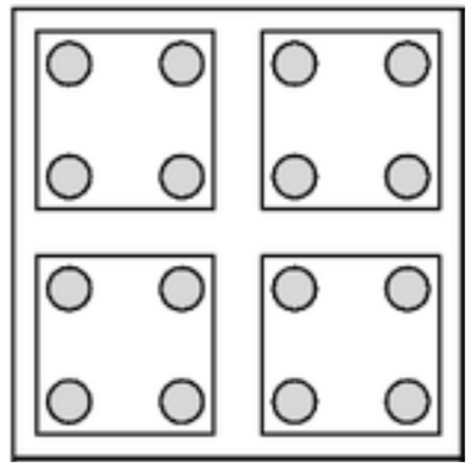
2. Reinforcement learning the transition kernel: Song et al, 1706.07561, Levy et al 1711.09268, Cusumano-Towner et al 1801.03612, Bojesen, 1808.09095



3. Performs MC in the variationally learned disentangled representation: Wavelet MC, Ismail 03' , NeuralRG 18'

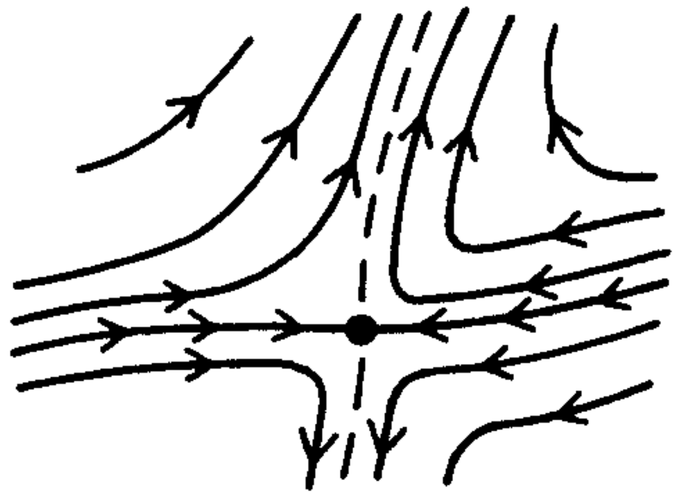


Deep learning and RG

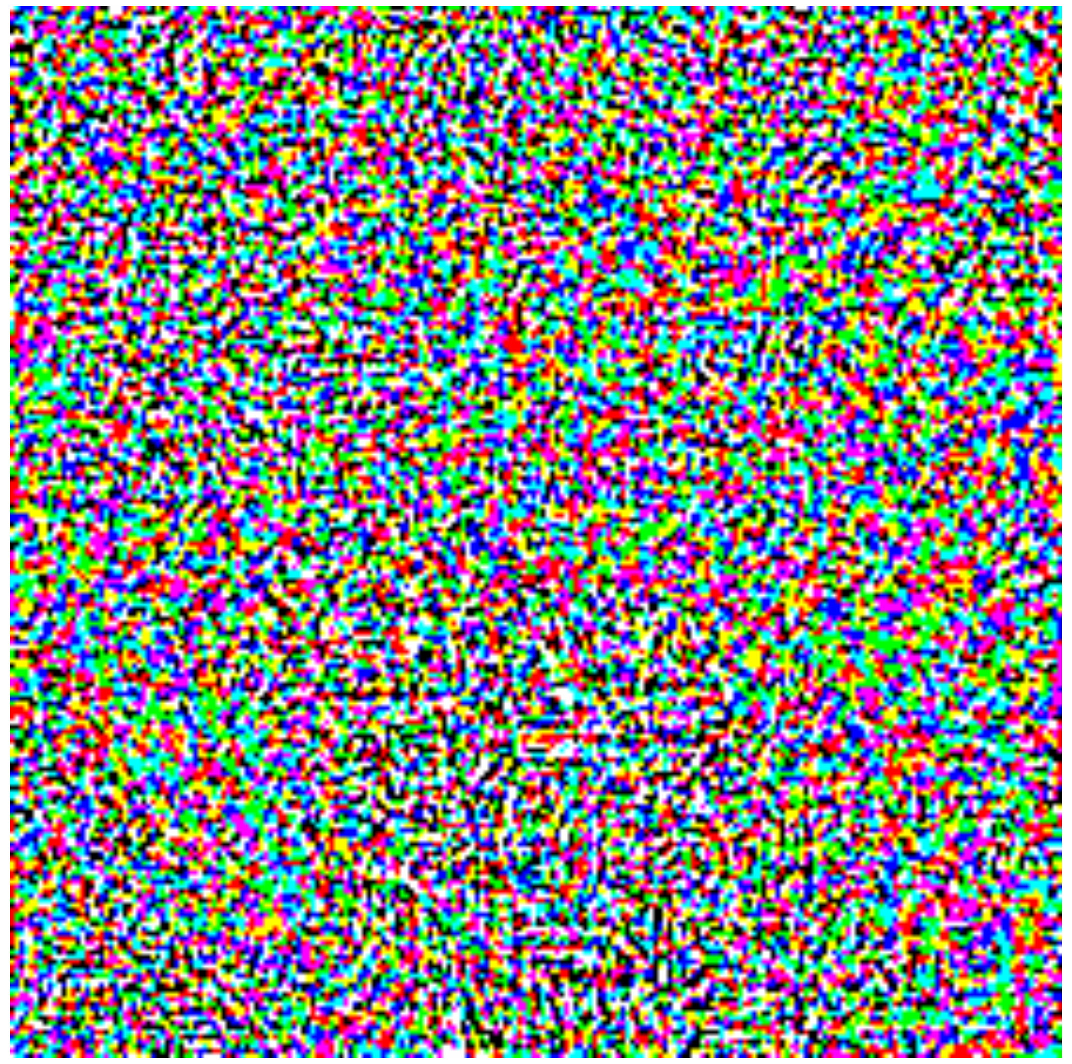


't Hooft, Gross, Wilczek, Kadanoff, Wilson, Fisher…

Bény, Mehta, Schwab, Lin, Tegmark, You, Qi …



$+ .007 \times$



$=$

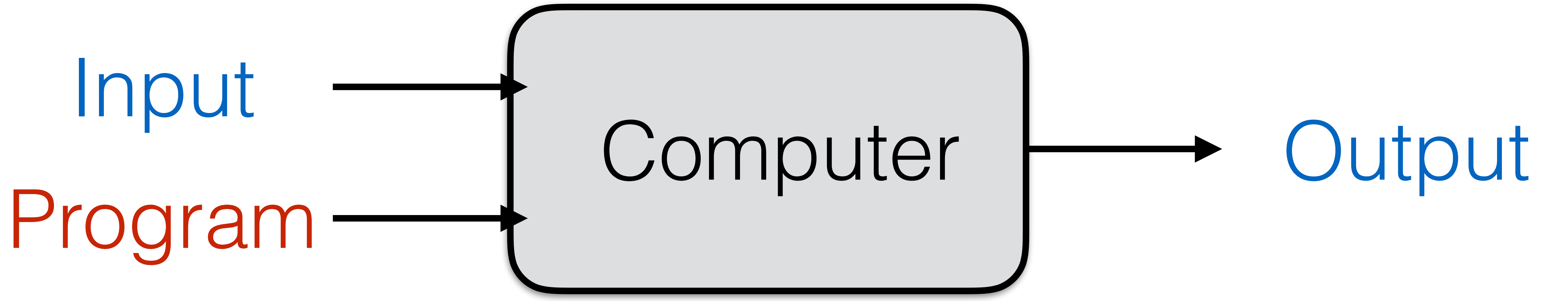
Panda
confidence58%

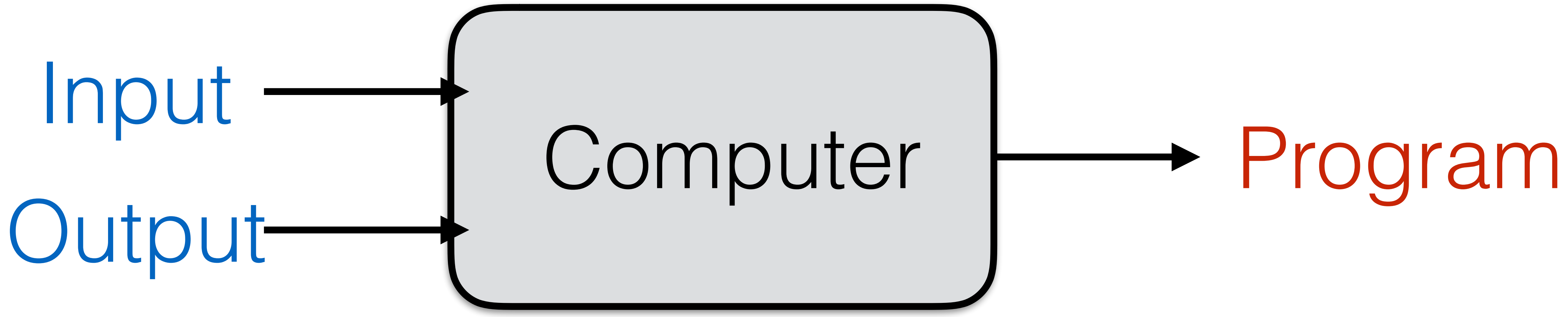
Goodfellow et al, 2014

Gibbon
confidence99%

Vulnerability of deep learning, Kenway, 1803.06111 & 1803.10995

Differentiable Programming

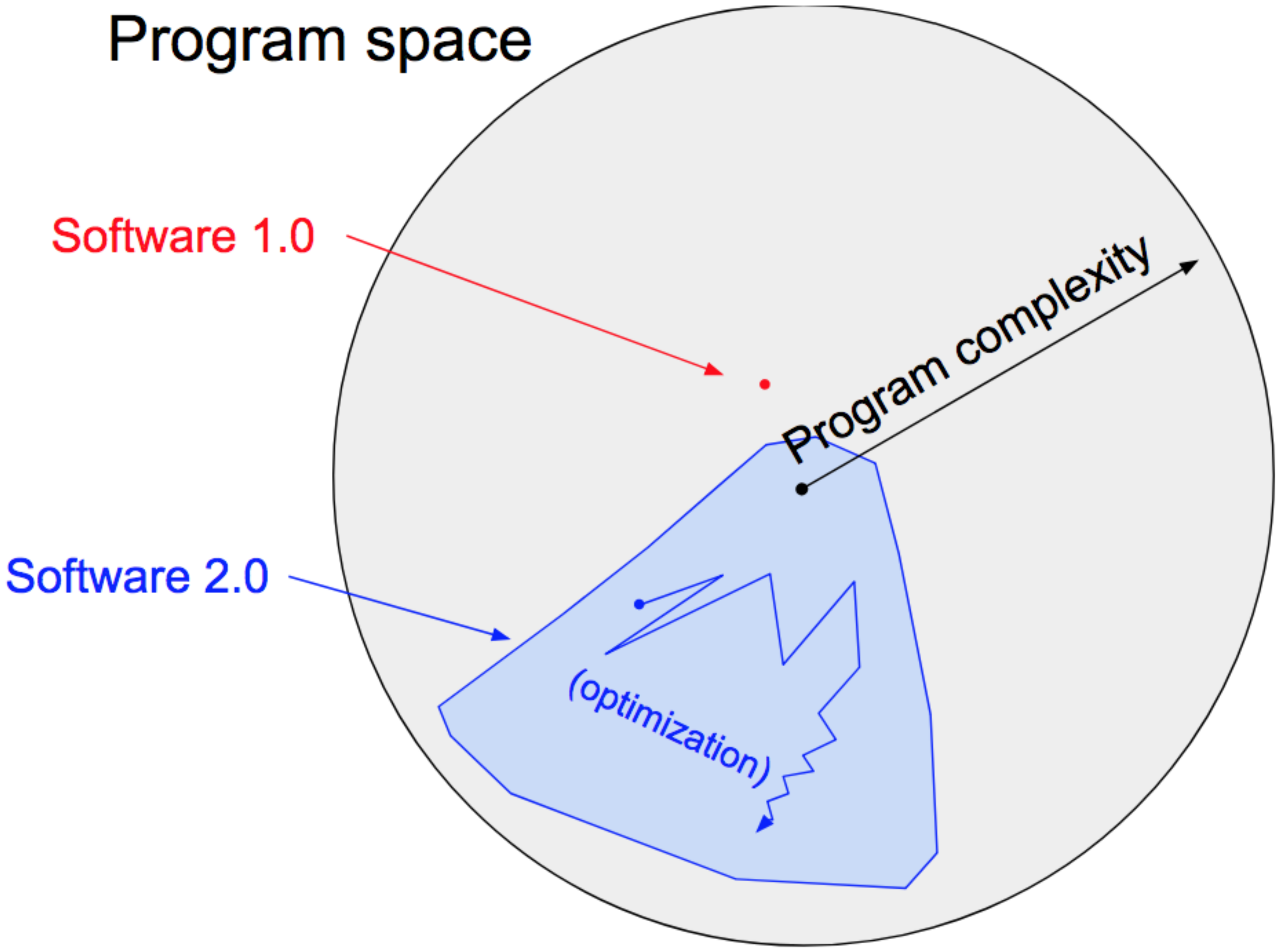




Andrej Karpathy

Director of AI at Tesla. Previously Research Scientist at OpenAI and PhD student at Stanford. I like to train deep neural nets on large datasets.

https://medium.com/@karpathy/software-2-0-a64152b37c35



Writing software 2.0 by searching in the program space

Differentiable Programming

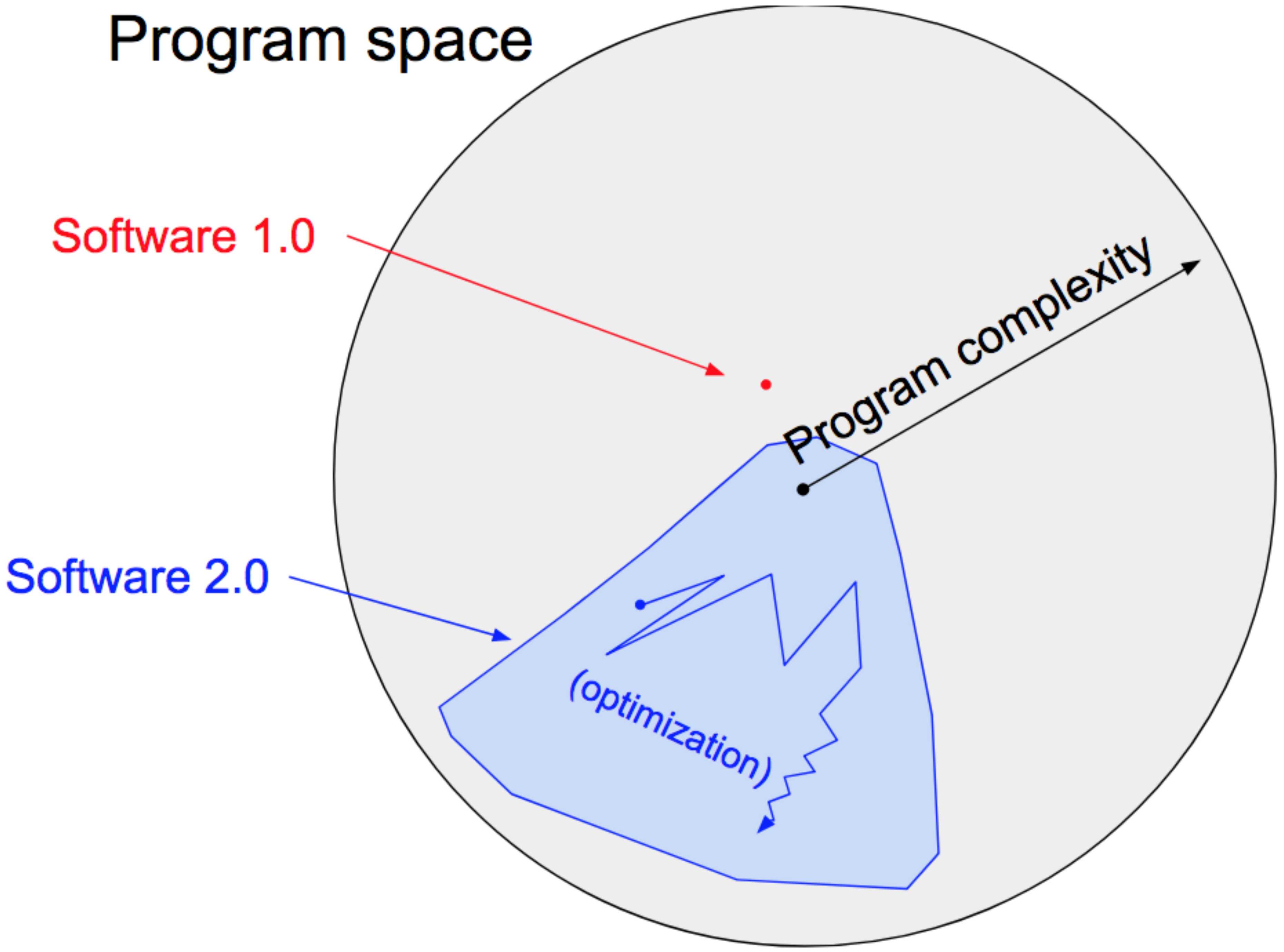
Benefits compared to 1.0

- Computationally homogeneous
- Simple to bake into silicon
- Constant running time
- Constant memory usage
- Highly portable & agile
- Modules can meld into an optimal whole
- Better than humans

Andrej Karpathy

Director of AI at Tesla. Previously Research Scientist at OpenAI and PhD student at Stanford. I like to train deep neural nets on large datasets.

https://medium.com/@karpathy/software-2-0-a64152b37c35



Writing software 2.0 by searching in the program space

Differentiable Scientific Programming

- Most linear algebra libraries are differentiable
- Condition/Sort/Permutations are also differentiable
- ODE integrators are differentiable with O(1) memory
- Differentiable ray tracer and Differentiable fluid simulations
- Differentiable Monte Carlo/Tensor Network/Functional RG/ Dynamical Mean Field Theory/Density Functional Theory…

Differentiable Eigensolver

$$AU = UD$$

Forward mode: What happens if $A = A + dA$? Perturbation theory

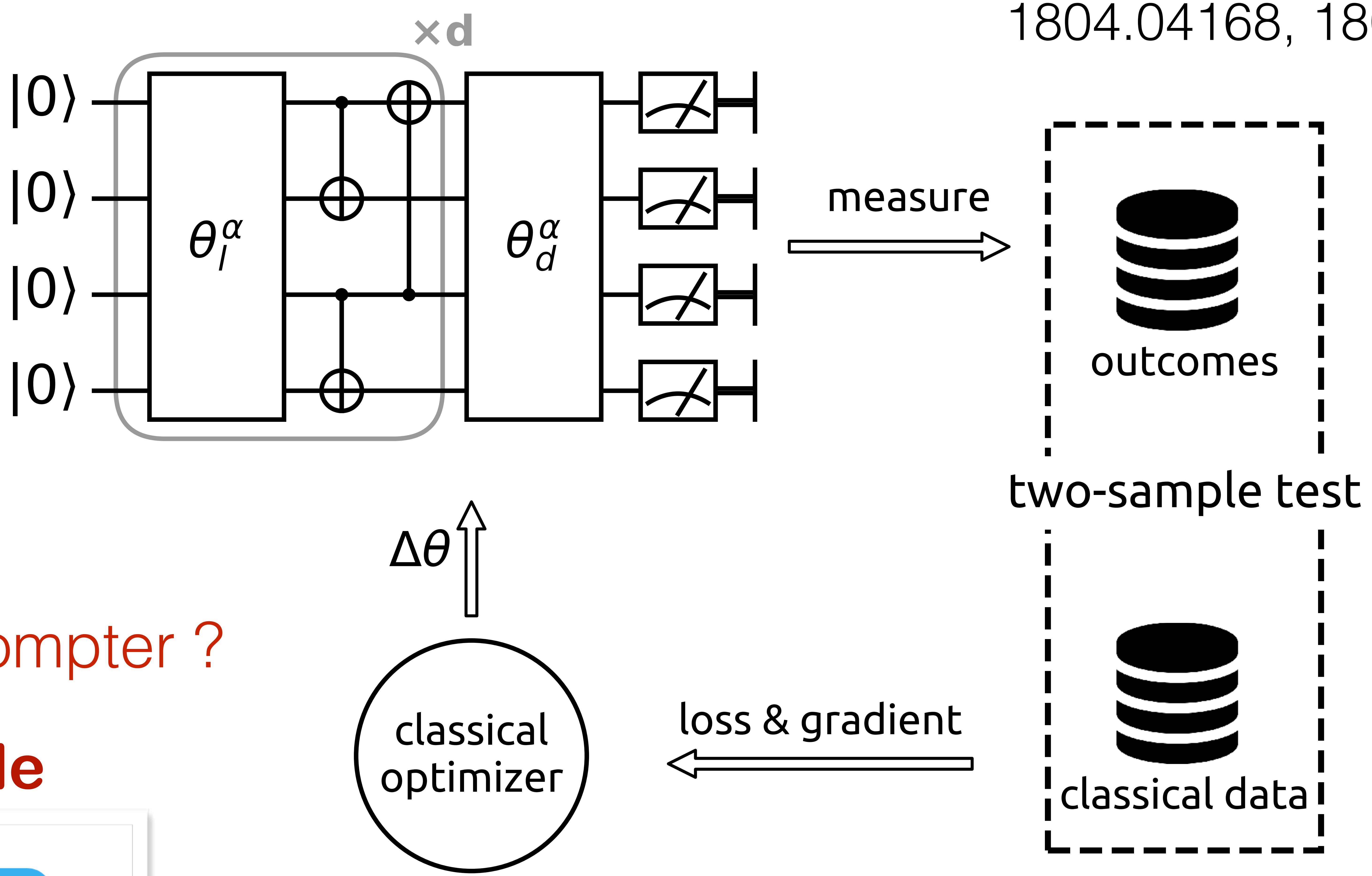
Reverse mode: How should I change A given dU and dD ? **Inverse perturbation theory!**

Hamiltonian engineering via differentiable programming

Differentiable Quantum Programming

With Liu, Zeng, Wu, Hu
1804.04168, 1808.03425

Short term:
What can we do with circuits of limited depth ?



Long term:
Are we really good at programing a quantum compter ?

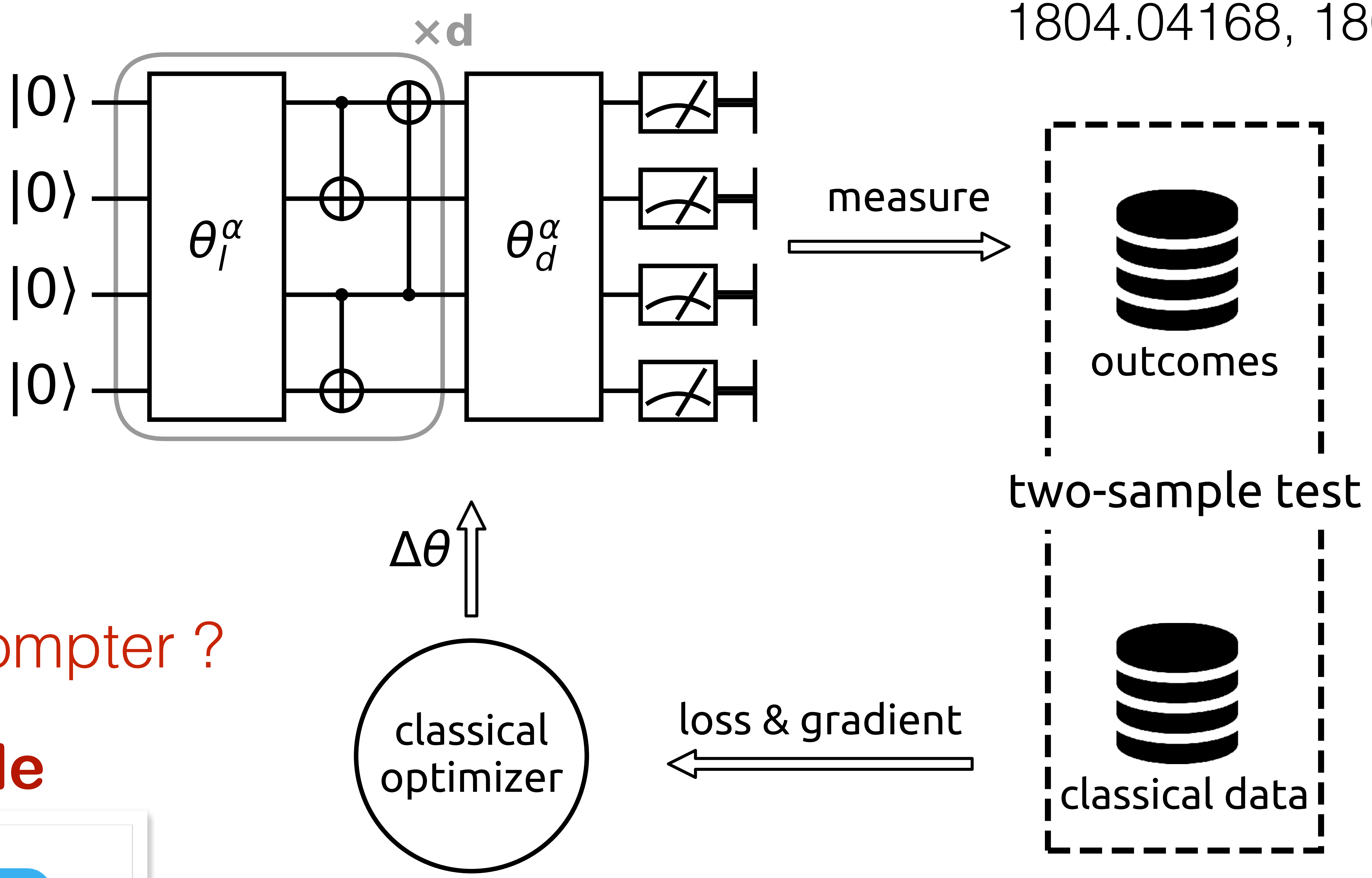
Quantum code

Differentiable Quantum Programming

With Liu, Zeng, Wu, Hu
1804.04168, 1808.03425

Short term:
What can we do with circuits of limited depth ?

Long term:
Are we really good at programing a quantum compter ?

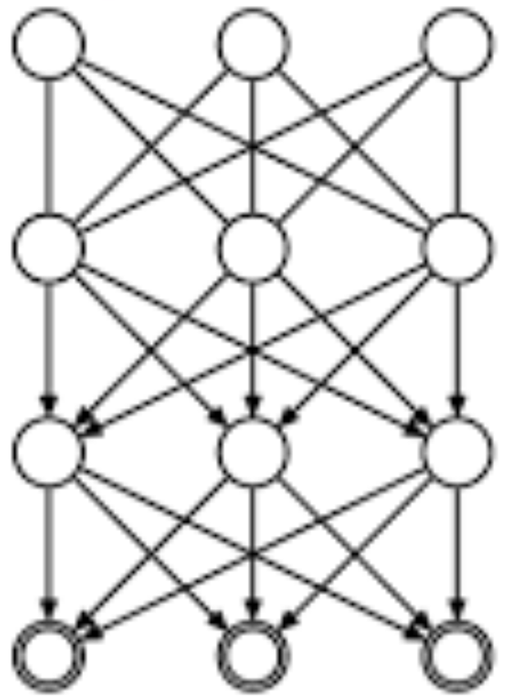


Quantum code

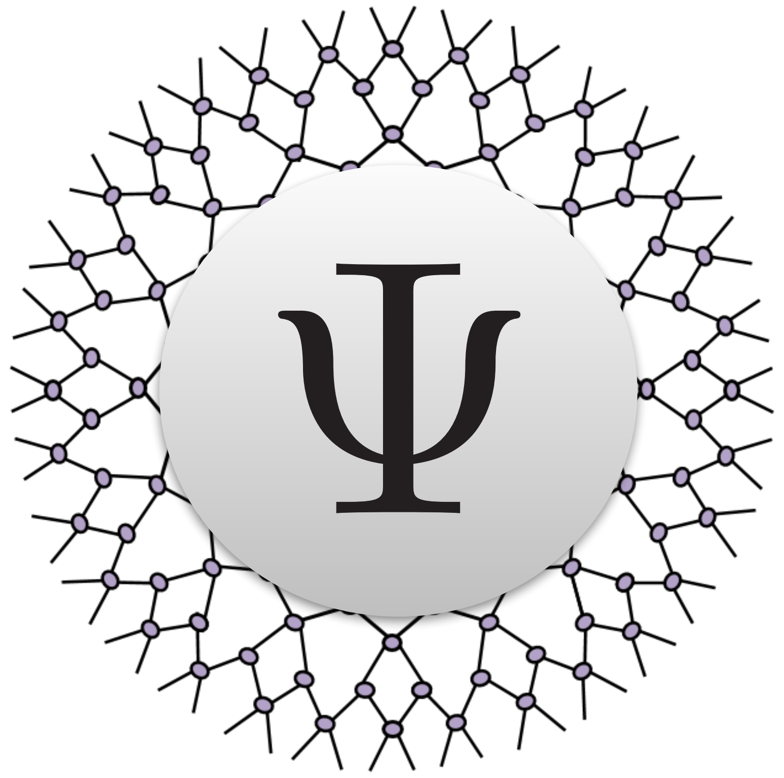
Xiu-Zhe Luo, Jinguo Liu
https://github.com/QuantumBFS/Yao.jl/

What is a deep neural network ?

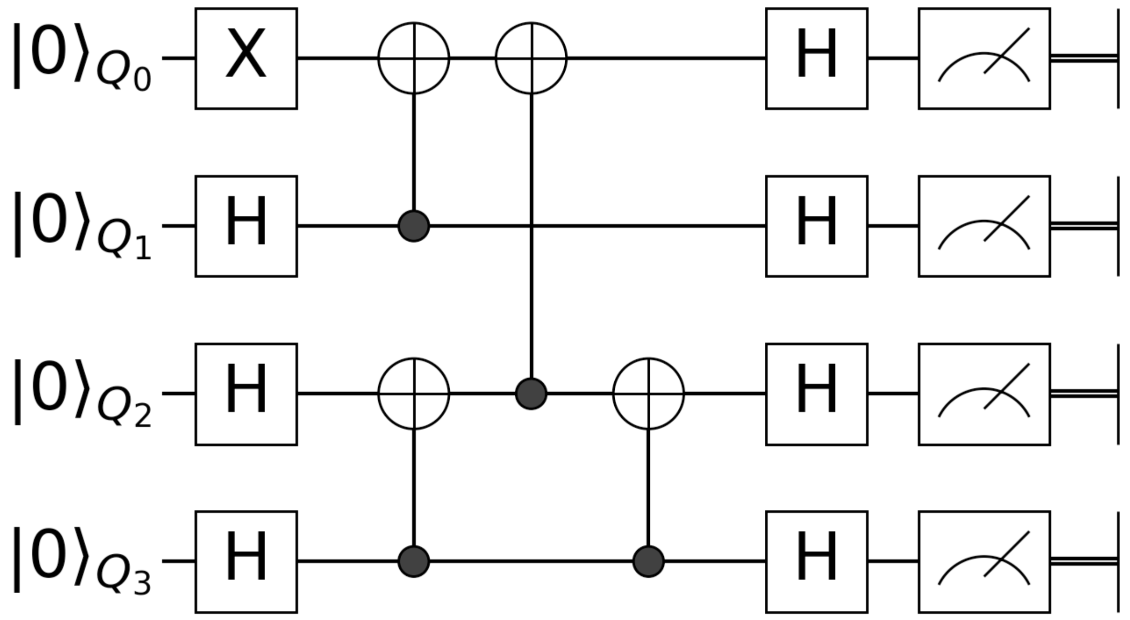
Neural Net



Tensor Net



Quantum Circuit



“三重境界”

1. Function Approximation
2. Probabilistic Transformation
3. Information Processing Device

Hands on time!

https://github.com/wangleiphy/dl4csrc

1. Back propagation from scratch
2. Differentiable Ising solver
3. Fun with normalizing flows

Thank You!

Jin-Guo Liu	Xiu-Zhe Luo	Pan Zhang	Song Cheng	Linfeng Zhang
Jinfeng Zeng	Yufeng Wu	Dian Wu	Shuo-Hui Li	Weinan E